Information seeking, gathering and review:
Journalism as a case study for the design of search and authoring systems

Simon Attfield

University College London

A thesis submitted for the degree of PhD

2004

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

It is often argued that research into information systems design must consider the complex relationships existing between information activities and the context of work or leisure within which they occur. Recent research has seen an increase in studies and models locating information seeking and use within its wider task context. Less common, though, are studies that extend this focus to the equally important issue of how users manage and manipulate found information in the service of a task. This thesis explores the relationship between writing tasks and information seeking and manipulation, using journalistic writing as a case study, in order to reason about the design of integrated information retrieval and authoring systems.

The empirical work begins with a lab-based exploratory study of news-writing which identifies phenomena for further consideration. Drawing on these, an interview study with journalists in a national newspaper newsroom provides a model of how journalists seek, gather and marshal information in the context of what is an uncertain and evolving task. Following this, writing and the way it relates to information behaviour is considered in more abstract terms based on the idea of writing as a design activity (Sharples, 1996). Drawing on research from design psychology, this perspective forms the basis for a conceptualisation of writing within which findings from the newsroom study and other key information seeking research findings (particularly uncertainty and focus refinement) can be understood and explained.

The thesis then draws together its various threads with a set of requirements for integrated information retrieval and authoring systems focussing once again on the task of journalistic writing. A novel experimental information retrieval and authoring system is described based on some of the requirements, and a study of journalists using the system is reported, validating theoretical contributions of the thesis as well as the system's functionality.

Table of Contents

CHAPTER 1 - Introduction

CHAPTER 2 – Literature Review

CHAPTER 3 – Orientation: A Lab-Based Exploratory Study

CHAPTER 4 – Uncertainty And Change: A Contextually Rich Model Of Newspaper Journalist's Information Behaviour

CHAPTER 5 – Conceptualising The Wider Problem: A Design Psychology Interpretation of Writing Tasks

CHAPTER 6 – Requirements And Design

CHAPTER 7 – Experimental Evaluation Of Aspects Of The Model And The Prototype

List of figures

List of tables

Acknowledgments

The work reported in this thesis was supported by the Engineering and Physical Sciences Research Council, and also by Microsoft Research Ltd, Cambridge, England.

I am very, very indebted to many people who have helped me while I have been researching and writing this thesis. Doing a PhD, I now know, is an intense personal journey, but for me, one in which key people have played an invaluable part. Whether it be by taking the time to engage with the work, by opening doors and providing resources, or by putting up with me being 'intense' for extended periods of time, these people have frequently helped me find the footholds for each next step and I would like to thank them here.

I would like to thank friends and colleagues in Computer Science at UCL for numerous stimulating discussions and useful leads. In particular, Sasha Brostoff, Paul Cairns, Maia Garau, Simone Stumpf, Michael Tscholl and Mischa Weiss-Lijn are well worth talking to. I would also like to thank my second supervisor Janet MacDonnell for her time, encouragement and helpful suggestions.

For proof-checking various chapters of this thesis I am very grateful to Yvonne Rose, Helen Wang and Jeff Wood.

I would also like to thank Steve Robertson of Microsoft Research Ltd. for his helpful thoughts and for making resources available to me.

I would like to thank the people who have helped me gain access to journalists to study, including Rod Allen, George Brock and Neil Maiden. I am especially grateful to Richard Dixon of The Times for championing my work there, helping me find my way around an unfamiliar organisation, and offering context to help me make sense of what I found.

I would like to express my warmest thanks to my supervisor John Dowell who has patiently advised me through many twists and turns and given his deep consideration and support to what I have produced on the way. I now frequently find myself sharing his many words of wisdom with others.

I would like to thank my mother, father and brother for their enormous encouragement and confidence, and I would especially like to thank my mother for her patient proof-checking. I also thank my son Louis and my daughter Zoë who have been extraordinarily patient with dad while he has been writing this thesis, and have at the same time been a well-needed source of fun and distraction.

Finally, I don't know how people do PhDs without Jess King. Jess has been endlessly patient and endlessly supportive. She has constantly made me feel that we were working towards a PhD together. The result is as much hers as it is mine.

Chapter 1

Introduction

The owner of the memex, let us say, is interested in the origin and properties of the bow and arrow. Specifically he is studying why the short Turkish bow was apparently superior to the English long bow in the skirmishes of the Crusades. He has dozens of possibly pertinent books and articles in his memex. First he runs through an encyclopaedia, finds an interesting but sketchy article, and leaves it projected. Next, in a history, he finds another pertinent item, and ties the two together. Thus he goes, building a trail of many items. Occasionally he inserts a comment of his own, either linking it into the main trail or joining it by a side trail to a particular item. When it becomes evident that the elastic properties of available materials had a great deal to do with the bow, he branches off on a side trail which takes him through textbooks on elasticity and tables of physical constants. He inserts a page of longhand analysis of his own. Thus he builds a trail of his interest through the maze of materials available to him.

(Vannevar Bush, 1945 p8)

1.1 Information behaviour research – the holistic perspective

In 1945, Vannevar Bush's essay 'As We May Think' appeared in *Atlantic Monthly.* It included a startling vision of the future of information storage and retrieval. Published 2 months before completion of the first general purpose digital computer[1] and 5 years before the first computerised information retrieval system[2], Bush's essay was an attempt to spark the imagination of scientists towards new technological possibilities that would help them meet the challenge of what was to be a post war information explosion. Bush proposed the idea of the Memex machine—a device that would store vast amounts of information; information that could then be instantly retrieved and linked into new customised trails (what he referred to as "associative indexing" (Bush, 1945, p.8)) alongside the user's own notes. Bush imagined the storage of books, pictures, newspapers, personal records and communications. He described documents projected onto translucent screens, many at a time; and a keyboard and levers for inputting commands and creating linked trails.

The Memex was partly a proposal for reducing the time required to find information—to make information-seeking from an ever increasing catalogue of available material more efficient. But more generally, the memex was visualised as a tool for enriching a user's general work situation. The perspective Bush took was from the point of view of the user and of the broader tasks that they might perform using information. The Memex was not simply an information retrieval tool, but also a tool for linking, creating and storing new information artefacts. As the Turkish bow scenario demonstrates, Bush did not consider information retrieval in isolation from a user's broader information activities but as deeply integrated within them. The memex, as Bush saw it, would be "the piece of furniture at which he [an individual] works" (Bush, 1945, p.7).

Bush's essay is often cited as a major inspiration behind information retrieval systems, hypertext and the Web. However, through increasing research specialisation and the inevitable reductive decomposition of problems into ever

[1] The first general-purpose programmable digital computer is commonly thought to be the ENIAC (Electronic Numerical Integrator And Computer), built at the Moore School of Engineering at the University of Pennsylvania in Philadelphia.
[2] IBM's Electronic Statistical Machine, Type 101.

smaller and more manageable parts, the vision of integration has become somewhat fragmented. Over the years, researchers in Information Science, and HCI for that matter, have isolated specific issues, including mechanisms and behaviours related to searching for and finding information, and have omitted to consider how these mechanisms and behaviours might connect within the integrated whole of information use in relation to real-life work tasks. Academic information retrieval research is often accused of treating the finding of information as an endpoint, to the exclusion of how that information might then be used within a wider set of task objectives. For example, in a 1999 review of information-seeking research, Wilson commented that,

> ...[in 1981] information use had received little attention and, within information science, that statement is still relatively true today.
>
> **(Wilson, 1999, p.251)**

And from a technological perspective, Kuhlthau and Tama recently noted,

> ...for the most part, information systems and services have been designed to support information seeking and gathering without consideration for accommodating the ultimate need of applying information to accomplish work tasks.
>
> **(Kuhlthau &Tama, 2001, p.26)**

Kuhlthau and Tama went on to argue that information systems should go beyond provision for seeking and gathering to support the interpretation and use of information as this occurs within the wider task context. In recent years, voices such as Wilson's, and Kuhlthau and Tama's have contributed to something of a sea-change within user-centred information-seeking research towards a recognition of the importance of studying information-seeking and information use within an overall task context. This thesis attempts to contribute to these efforts by exploring information behaviour and the design of information retrieval and authoring tools for users in the domain of journalistic writing.

Wilson has defined 'information behaviour' as

> ... the totality of human behavior in relation to sources and channels of information, including both, active and passive information seeking, and information use.
>
> **(Wilson, 2000, p.49)**

The aim of this thesis is to provide an account of information behaviour within a complex information task in order to explore the design of information tools and the development of information behaviour theory (including information-seeking and its subsequent manipulation and use). In Wilson's definition, whilst information use is included within 'information behaviour', it is perhaps played down somewhat, almost as a minor caveat. In this thesis 'information behaviour' is taken in its widest sense, including not only of what information is sought and how it is sought, but additionally how it is gathered, reviewed, perhaps incorporated, perhaps discarded, and in any other ways worked, manipulated and potentially put to use.

Specifically, the focus is on information behaviour in the context of writing, taking news report and feature article writing by newspaper journalists as a case study. In the first instance, the concern is with understanding how a task, rich in information-seeking, shapes embedded information behaviour and how this behaviour relates to and shapes writing. In the second instance, it is concerned with exploring the implications that this understanding has for the design of integrated information systems and for the development of general information behaviour theory.

Information-seeking and associated information behaviour may occur in many contexts, and in support of many different kinds of task—a doctor may search medical databases to help them make a clinical decision, a lawyer may consult case records to help them construct a defence. The value of studying information behaviour in the context of a writing task (*i.e.* document authoring) is that writing in particular so often generates a requirement for information. And the reason for exploring journalism is that it provides a particularly suitable focus for the study of information behaviour in a high-tech work context. Writing a news reports or feature article motivates a great deal of information behaviour, yet little is known about what this behaviour is and how it relates to its wider task context. Journalists make extensive use of electronic information sources, including *electronic news cuttings* (ENC) services, and modern journalism with its fast pace and need for competitiveness places high demands on the usability and efficiency of supporting information systems.

To some extent, perhaps, adopting such a particular focus may mean that findings from the thesis are tied to the domain of journalism. However, all

research has to adopt some focus and findings may be generalisable elsewhere. Importantly, by understanding *why* findings occur and also through comparison with research from other task domains, generalisation is possible.

Over the past 20 to 30 years journalism has changed significantly. Today's newspaper reporter is no longer the shadowy gumshoe hack of the 1930's novel chasing down leads armed with only a notebook and a good nose for a story. Today, they have computers. Modern-day journalists spend much of their time in brightly lit offices in front of workstations which they use to monitor breaking newswire alerts from Reuters or The Press Association, search ENC archives storing millions of articles from local, national and international sources, print documents, and manage and store extracts of text in new documents. They scour the Internet for contact details, tap out reports on a keyboard, and file their copy electronically to be picked up by a sub-editor perhaps sitting a few desks away. Some newspaper articles don't even materialise as ink on paper, instead being uploaded onto newspaper websites. Through electronic document archives, search engines, word processors and the World Wide Web, much of Bush's vision is now reality, and the work of the journalist, as with work in nearly all walks of life, has been transformed.

The electronic revolution in news organisations is relatively recent and is ongoing. Until quite recently a major asset of any news organisation would have been its library of categorised newspaper cuttings—files and files of classified paper clips that were continually being updated by dedicated library staff. But then ENC services were introduced in the 1980's. In the US newspaper industry 1985 was a "watershed year" for the installation of ENC databases with the advantages beginning to outweigh the costs (Semonche, 1993). Although at this time, systems tended to be confined to library departments with searching normally performed by librarians.

During the early 90's uptake of ENC services in media organisations increased considerably, and searching by journalists became much more commonplace (Martin & Nicholas, 1993). Access to full-text news databases, such as NEXIS, VU/TEXT, Datatimes, Dow Jones News Retrieval System, BRS and DIALOG became more frequent in larger news libraries (Semonche, 1993). In a large survey carried out in 1991, Hansen and Ward found that, of 105 newspapers

with circulations over 100,000, 67% had an electronic library and 90% had commercial database subscriptions (Hansen & Ward, 1991).

By the late 90s, access to online news cuttings archives had reached the journalist's workstation in many news companies. The BBC and the Associated Press are now leading the news media software market with a system called ENPS (Electronic News Production System)—a fully integrated newsroom system incorporating multiple information services including ENC information retrieval tools and text editors within a single software environment.

The job of the journalist is to inform, to question, to educate and to engage the reader. Sometimes it is to let people tell their story or to act as watchdog. Sometimes it is to entertain, but always to present information as accurately and objectively as possible and to help people understand what a story means for them. Technologists designing for the news-media industry have many opportunities with which to support the journalist as an effective gatherer and conveyor of a particular kind of information. But technology to support information-seeking and use applications cannot simply be 'thrown' at a work situation without an appreciation of the specific needs of the situation, and this means understanding how people work and how they would ideally work.

Two related issues can be considered when addressing the question of improving the relationship between the work of the journalist and the technology they use. First, the kinds of information-seeking and use tools most frequently used by journalists are designed along relatively generic lines with minimum tailoring for their particular kind of work. Commercial ENC services are accessed via client-side browsers. Their search engines perform full-text keyword searching over newspaper and magazine article archives and present their results in much the same way as any other information retrieval system; it is for the journalist user to adapt the way they work to this general purpose functionality. For example, searching for an interview with a particular film director may be a question of performing a search using the director's name and then laboriously browsing an extensive results list. The adaptation of generic tools will inevitably be less efficient than the use of tools tailored to the task at hand. A close examination of journalists' work provides a vantage point from which to consider the design of tools to dovetail more closely with their work.

The second issue, which is perhaps a corollary of the first, relates to the separation between the various software tools that journalists currently use, including information retrieval tools and the tools that they use to gather and manipulate information and generate new documents. Developers, like academic researchers tend to specialise. Perhaps for reasons of expertise or commercial advantage they tend to develop tools to support one kind of activity across a range of task domains. Hence, in any task situation, people usually find themselves using a combination of separate, and more or less generic software tools. But seeking, manipulating and writing information are not discrete, unrelated behaviours, but form parts of a single, continuous flow of interdependent activity, each constantly modifying and conditioning the other. So, by reducing the separation between tools—by integrating them—opportunities should arise for interrelating actions across them and for better enabling natural continuity. Such interconnection has not been explored within the ENPS system, although its commercial success in recent years perhaps offers an argument for integration, and a point of departure from which to consider how activities, hitherto viewed as independent, might be considered together and more tightly integrated around a single task objective.

1.2 Three research Questions

The research questions for this thesis are:

1. **What are newspaper journalists' prototypical information behaviours in relation to the seeking and use of information from electronic news cuttings services, whilst writing news reports and feature articles, *and* what are the aspects of their task situation that explains them?**

   Question 1 is a composite question—although its parts are closely linked. It asks what are the prototypical information behaviours of journalists writing news and feature articles in relation to the seeking and use of information from ENC services. The term 'prototypical information behaviours' refers to generalisable descriptions of what information is characteristically sought and how it is sought, and of the ways in which information is gathered, reviewed, incorporated, discarded, and in any other way worked. Information is regarded as a basic raw material from

which new information artefacts (*i.e.* news reports and feature articles) are shaped.

The particular focus taken by the question is on writing news reports and feature articles. Although there may be information activities which journalists perform outside of a given writing assignment, such as generally keeping up with current news, the particular scope taken by the question is on activities which occur specifically as part of the process of writing an article for publication, and in particular, where that article is a news report or feature.

In addition, the perspective taken is one of the individual newspaper reporter/writer and their activities covering the period from when they are briefed by their editor on their assignment to the point at which their finished copy is filed. Although news writing can be a collaborative activity, particularly on big stories, usually it is not. For the most part authoring a news or feature article can be regarded as an activity of a single individual. Once a journalist files their copy, however, things are different since then it will be reviewed and potentially modified by sub-editors before the newspaper finally goes to print. Although, the scope taken by research question 1 is of the work of the individual author *within* this broader process, it is not of that broader process *per se*.

The question also defines as its scope a concern with information behaviour as it occurs in relation to electronic news cuttings services. Information is the principal commodity for the journalist and is sought through dozens of sources including newswire feeds, email, written reports, interviews and discussions with colleagues. But, one of the most significant resources at the journalist's disposal is the online cuttings archive, and, as recent commercial developments in the news-media industry have shown, this resource lends itself particularly well to digitisation and automated search by journalists as end users.

The second part of the question states the aim of locating the object of enquiry stated in the first part within the context of those aspects of the task situation which can offer explanatory leverage. Hence, the aim is not just to report information behaviour but to attempt to explain why it

happens and why it happens the way it does. Of particular concern here are two issues. First, there are the goals of news and feature writing and how it's wider influences, whether these have their root in social, political, legal, commercial, linguistic, or other external issues, propagate down to shape the writing task and in turn shape its embedded information behaviour. Second, there are the *means* by which journalists meet their goals which undoubtedly shape what is done and how it is done. In this latter category is included not only the tools with which journalists work but also the journalists themselves as cognitive agents operating within a complex work environment, managing and progressing their work through a sequence of assessments and choices.

The general trend towards providing accounts of information behaviour as located within a wider context has been discussed briefly. Here a holistic, contextualised perspective is motivated by the views that:

- We have an explanation for something once we understand its place in the whole. As with many complex activities, journalism is a process existing within and in virtue of a system; without it the journalist's information behaviour, indeed journalism itself wouldn't exist. This natural setting plays a fundamental role in creating and conditioning the journalist's task and how it is achieved. Consequently, journalists' information behaviour is most completely explained with reference to that system and journalists as agents acting within it;
- Identifying contextual dependencies can delimit the scope of generalisation. First, explaining the object of research (in this case a particular kind of behaviour) contributes to the credibility of that behaviour as representative of the situation under study. Second, understanding the context for behaviour can suggest why comparable situations might lead to different or similar outcomes;

2. **How does this knowledge relate and contribute to more generalisable theory of information behaviour in relation to the processes and structure of complex information tasks?**

The emphasis in Information Science on obtaining rich models of information behaviour as a function of its ecological context is predicated on the idea that since information behaviour depends so much on wider contextual issues (such as motivating task, organisational culture etc.) these must be considered if we are to fully 'understand' what's going on. But in this thesis the holistic perspective implied by question 1 and set out in the subsequent elaboration is not intended to preclude the possibility and the value of abstracted, generalisable theory.

As an abstraction from particular observations, theory permits research findings to be related to and evaluated in terms of new situations from those observed. Hence it enables research communities to collaborate around and develop common sets of ideas, providing coherence to what might otherwise be disparate and distributed research efforts. Indeed, the motivation for holism as laid out in the final point of the elaboration pointed implicitly to the value of abstraction as a means of generalising common themes between situations. And only by many people studying many situations can common themes be identified and disassociated from the contextually specific.

This is a grounded, inductive approach to developing generalisable theory according to which specific findings can be related to each other and discussed in more general terms. As such, it corresponds with the 'alternative' approach to studying purposeful action that Suchman (1987) identified as having emerged principally from Anthropology and Sociology, insofar as it recommends "... building generalisations inductively from records of particular, naturally occurring activities, and maintaining the theory's accountability to that evidence." (Suchman, 1987, p179).

It will be seen in chapter 2 that the development of a common body of theory has been of particular concern to the Information Science community for a number of years. As a direct effect of this concern, a

number of theoretical perspectives have emerged, and in Chapter 2 some of the more important theories in Information Science as well as Human Computer Interaction (HCI) are discussed. Particular emphasis is placed on general theory and theory relating to journalists' information behaviour that help explain observations made by the thesis, and also which can be developed in the light of those observations.

Question 2 articulates the goal of contributing to a more general understanding of information behaviour in relation to the process and structure of complex information tasks. In part, this means relating research findings to, and in some places extending existing theories. In part, it also means developing new conceptualisations which are generalisable across task domains beyond the immediate subject of study here (i.e. journalistic writing). The question also draws attention to two complementary themes: task process and task structure. The first theme is concerned with what people do—how they address and solve problems. The second theme is concerned with the structure of those problems *i.e.* what the problems are. Both of these themes are addressed by the thesis.

3. **What are the implications of the findings for integrated information retrieval and authoring systems for use by journalists (and others)?**

The aims and subject matter of this thesis lie at an intersection of Information Science and Human Computer Interaction. To a large extent, the driving force underlying research in both of these areas is with developing knowledge for the design of better tools or systems. In the case of Information Science, the concern is with systems that "facilitate the effective communication of desired information between human generator and human user" (Ingwersen, 1992, p.49). HCI, on the other hand, focuses on the design of technological systems from the point of view of promoting more effective or satisfying interaction, but, within this scope, extends its interests to include any kind of activity in which people interact with technology.

The motivation for researching information behaviour in journalistic writing in this thesis is to inform the design of systems for the retrieval and use of information that can enable better and more productive

interaction. Research can influence practice in more or less direct ways. In considering the implications of the work in the areas of Information Science and HCI, research question 2 can be considered as indirectly concerned with developing better design knowledge. The more we know about how people come to need, find and use information, and the better we understand the problems that motivate this behaviour, the better we ought to be able to design systems to help them. Question 3, however, is more directly concerned with exploring and validating design implications of the empirical findings of the thesis. Within the scope of this question are the generation of requirements, system design and building, and, ultimately, evaluation. To a large extent this scope will relate to systems intended for journalists, and in particular systems that integrate information retrieval services and text authoring tools within a single software environment. But, depending on the generalisability of the findings into other work domains, the scope may also extend to other kinds of writing task where those tasks also feature a proportion of research from online information resources.

1.3 Contributions of the thesis

The contributions of this thesis are:

1. A model of journalists' information behaviours in the context of the wider task of writing news reports and feature articles;
2. A general framework for representing writing tasks which accounts for a number of key information-seeking phenomena;
3. A set of design requirements for integrated information systems for journalists;
4. The design of a prototype system that is sympathetic to users' evolving interests as a function of their developing task focus;

1.4 The remaining chapters

Figure 1.1 summarises the structure of the thesis in terms of each of the remaining chapters. The figure describes the role of each chapter, where each of the main contributions appears, and where each of the three research questions (RQ1, RQ2 and RQ3) are addressed. Also shown, are arrows

representing how each chapter leads on from a previous chapter. These arrows show that after research question 1 (and to some extent research question 2) have been considered in the early chapters, the flow of the thesis splits with chapter 5 focusing more directly on research question 2, and chapters 6 and 7 focusing on research question 3. These separate streams are then brought together in chapter 8.

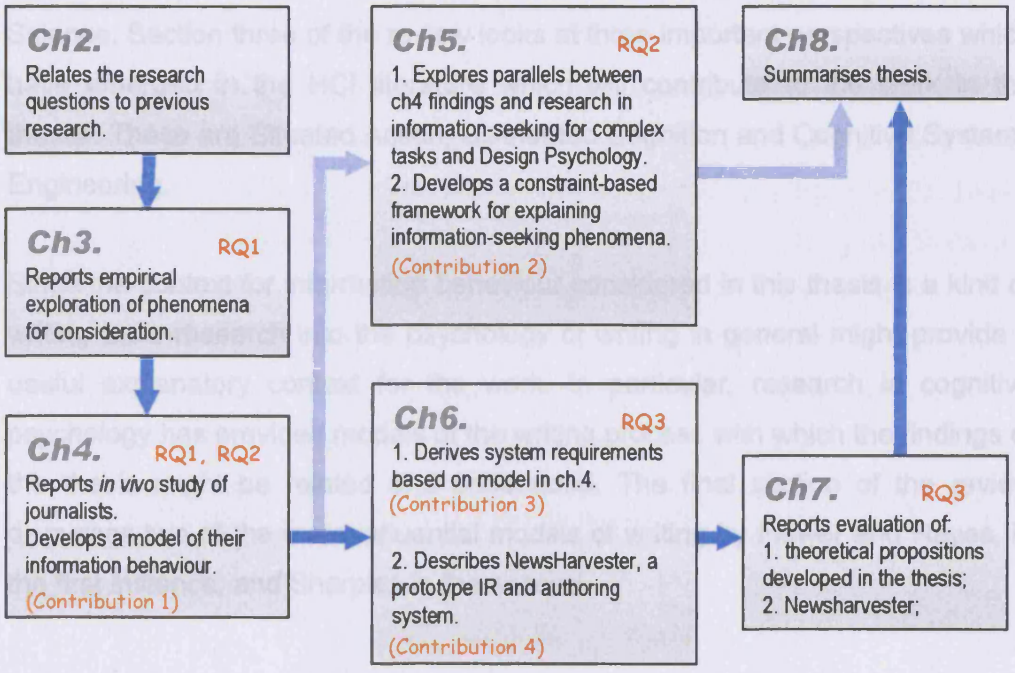


Figure 1.1 The structure of the thesis in terms of the remaining chapters.

1.4.1 Chapter 2 – Literature review

The role of chapter 2 is to locate the research questions within a body of previous research in the areas of Information Science and HCI, and to some extent Cognitive Psychology. The review is divided into three main sections. The first and largest section, tracks aspects of information-seeking research in Information Science which are particularly relevant to the thesis and includes influential metatheoretical approaches and mid-range models. The themes developed in this section are: the study of information-seeking in a task context, information-seeking as part of an evolving process, information-seeking uncertainty, the drive for the development of theory, cognition and the Cognitive Viewpoint, the Sense-making approach and the Behavioural approach.

The second section of the review covers Information Science research with a particular focus on the work of journalists, and in particular, work by Dave Nicholas and colleagues.

The role of context in shaping and determining behaviour, and therefore as an important object of behavioural research, has become a theme increasingly emphasised in HCI as it has been in information-seeking research in Information Science. Section three of the review looks at three important perspectives which have emerged in the HCI literature which will contribute to the work in the theses. These are Situated Action, Distributed Cognition and Cognitive Systems Engineering.

Since the context for information behaviour considered in this thesis is a kind of writing task, research into the psychology of writing in general might provide a useful explanatory context for the work. In particular, research in cognitive psychology has provided models of the writing process with which the findings of the thesis might be related and understood. The final section of the review discusses two of the more influential models of writing by Flower and Hayes, in the first instance, and Sharples in the second.

1.4.2 Chapter 3 – Orientation: A lab-based exploratory study

The empirical work of the thesis, most directly addressing research question 1, begins with a lab-based exploratory study of news writing reported in chapter 3. The purpose of this first study was to provide some orientation for the remaining work with an exploratory, data-driven identification of behaviours and related phenomena for further consideration. The study was a lab-based, qualitative, observational study of non-journalist subjects performing a task which simulated a news research and writing scenario. Although there has been research which has explored information needs in journalism and some aspects of their search behaviour (Nicholas & Martin, 1997), and also journalists' use of the internet (Nicholas, *et al* 2000), there has not been a systematic consideration of the full range of journalists' information behaviours during writing assignments and the ways in which these relate and integrate within the process of research and

writing. The aim was to draw attention to information behaviour which might provide some focus for a subsequent field study (reported in chapter 4).

The lab-study data (screen recordings and associated talk-aloud voice recordings), is analysed at the level of global patterns throughout the task and also at the level of more local phenomena.

1.4.3 Chapter 4 – A study in a newsroom

Drawing through some of the findings of the lab-study for focus, chapter 4 addresses research questions 1 and 2 by reporting an *in vivo* interview study conducted with journalists at *The Times* in London, developing a model based on the findings, and also relating the findings to other models in the literature. Using the Grounded Theory approach (Strauss & Corbin, 1990 & 1998), which in the study is additionally structured by concepts taken from The Cognitive Systems Engineering (CSE) framework (Rasmussen, Pejterson and Goodstein, 1994), an ecologically grounded model of newspaper journalists' information behaviour is developed which relates the use of information from online ENC archives during the process of researching and writing news report and feature articles.

The contribution of CSE to the study is to view complex, discretionary work as occurring within a context of constraints and resources. Hence, these two kinds of element provide the study with a notion of 'context' on which to base explanatory accounts of user activity. Constraints demarcate the space of potential solutions into more or less acceptable outcomes and so determine what work is done; resources are utilised in order to meet or optimise the active task constraints and so determine how work is done (within the model the user's cognitive faculties are cast as a kind of resource).

In accordance with the framework, the model describes the discovered information activities as its central feature and located within explanatory constraints and resources which emerged from the data. At their top level the information activities decompose into *information-seeking*, *information-gathering* and *information reviewing*. Importantly, the model describes journalistic research and writing as a dynamic and uncertain task as a result of frequent mid-assignment constraint changes and the journalist's own evolving concept of the finished product. As new information is encountered and as events unfold, new

insights are gained and so new facts and issues can become important. As a consequence of this uncertainty, information relevance judgments can change and so commitments to information usefulness, whether these are positive or negative, are only ever provisional.

1.4.4 Chapter 5 – Conceptualising the wider problem: a design psychology interpretation of writing tasks

Chapter 5 is a theoretical chapter most directly focussed on research question 2. Here the task scope is broadened beyond journalistic news and feature writing to a more general view of writing as a common context for information behaviour.

Chapter 5 is divided into two parts. First, adopting the view, most comprehensively articulated by Sharples (1996), that writing is a class of design activity, the chapter sets out to identify and explore parallels between findings from the field study, other findings from research into information-seeking in the context of complex task performance, and ideas that have emerged from research into the psychology of design. To provide context, the first section begins by summarising some of the information-seeking research reviewed in chapter 2 which will be useful for making the case for parallels in design psychology. Important themes in this work are uncertainty, its relationship with the formulation of a task focus, and the effect that this has on relevance judgments and query specificity.

The remainder of the first part of the chapter is then organised in terms of four features that design psychologists have identified as characterising design problems and design problem solving: incomplete specification, primary generators, the analysis/synthesis dynamic, and multiple constraints and integrated solutions. In each case the feature is described, related to information behaviour research, and exemplified by findings from the field study of journalists reported in chapter 4. In particular, the perspective provides a conceptualisation of writing within which findings from the newsroom study and other key information-seeking research findings (particularly uncertainty and focus refinement) can be understood and explained. It is shown that parallels and explanatory leverage can be found in what, for the information-seeking community, is perhaps an unfamiliar branch of research.

In the second section of chapter 5 the design perspective is used as a basis for a constraint-based framework for the representation of writing tasks which it is claimed offers a new explanatory framework for interpreting information-seeking phenomena. Like Sharples (1996), the framework represents a complex problem, such as writing, as the combination of its constraints taking the problem as viewed from the user's perspective. Sharples analysis, however, only considered information-seeking in passing. The framework in chapter 5 offers a constraints-based perspective that explains how writing can give rise to information needs, and, reciprocally, how found information can re-structure the task. The framework is elaborated using two journalistic writing/information-seeking scenarios as examples. Within these scenarios, vague, poorly specified information needs are represented as well as a well-specified and precise information need. Together the scenarios demonstrate how the framework can express the progressive refinement of focus during a writing task and how this relates to information-seeking behaviour.

1.4.5 Chapter 6 – Requirements and design

In chapter 6 attention turns to research question 3 and the implications of the findings of the thesis for the design of integrated information systems for journalists.

Following a short discussion of the nature of requirements and an overview of some electronic information-seeking technologies, chapter 6 returns to the information activities reported by the field study in chapter 4, and, dealing with each in turn, derives a set of system requirements. Included with each requirement is a short discussion outlining how it might be achieved in technological terms.

Also in chapter 6 one of these requirements—that systems should maintain connections or 'threads' between copy-and-pasted extracts and their source documents—is used as the basis for the design of a prototype integrated information retrieval and authoring system called NewsHarvester. This system, which is described at the end of chapter 6, incorporates information retrieval search, results display and document display with a text editor for gathering text extracts and writing copy. The system supports drag-and-drop functionality

between the document display and the text editor and incorporates the 'threads' idea by the automatic suffixing of extracted text with hyperlinks which redisplay its source document (Autolinks).

1.4.6 Chapter 7– Experimental evaluation of aspects of the model and the prototype

In chapter 7 the final study of the thesis is reported. In this study NewsHarvester was used to evaluate a series of theoretical propositions as well as the Autolinks functionality in a comparison with gathering information using standard-drag and-drop, and printing documents and optionally highlighting or annotating.

Chapter 7 describes three variations of NewsHarvester that were developed, each featuring one of the three information-gathering methods. Each design variation was used as a condition in a repeated-measures, cross-over design with counterbalancing between tasks and conditions. The Autolinks functionality corresponded with the experimental condition with the other two gathering methods used as reference conditions.

In the study, 15 journalism students, each with some professional experience, were asked to write news reports using each version of NewsHarvester. Their tasks were to research and write news reports on the basis of invented, but credible newswire reports and a verbal editor's 'brief'. The tasks had previously been validated (and in some cases modified) as representative of assignments tackled by newspaper journalists on a daily basis by an Executive Editor at *The Times*.

The results of post-task questionnaires and objective measures are reported along with their analysis. This provides validation of the theoretical propositions and a qualified validation of Autolinks functionality in comparison with the reference alternatives used in the study.

1.4.7 Chapter 8 – Discussion

In chapter 8, the outcomes of the research reported in the thesis are summarised in terms of the research questions posed in this chapter.

Chapter 2

Literature review

2.1 Introduction

The aim of this thesis is to explore information behaviour in the context of writing tasks, with journalistic writing taken as a case study—and this in order to reason about the design of integrated information retrieval (IR) and authoring systems and to contribute to theory. In this chapter, selected previous research is reviewed with the aim of providing a perspective within which to locate the research questions for the current work and the adopted meta-theoretical approach, and also to provide reference points for discussing issues through the course of the thesis.

Empirical research progresses through an interplay between empirical findings, models and meta-theoretical commitments and these lie on a scale of varying abstraction. Empirical findings derive directly from observations made of the objects of research. Models, whether these are inferred from or imply findings, serve to organise findings into more generalised structures. Finally, meta-theoretical perspectives are concerned with what sorts of things the objects of research should be, and, methodologically, how claims about them can be made. Since previous research can (and should) influence a research programme at the level of findings, models and meta-theory, contributions at all levels are discussed in the review. Further, since these are typically interdependent and interrelate closely within research activity, for the most part all three are integrated within a single review narrative.

The subjects covered in the review are: information-seeking and information behaviour in general, information-seeking in journalism, perspectives in Human Computer Interaction (HCI), and to some extent, the psychology of writing. In chapter 5 ideas from the psychology of design will also become important, but these will be introduced in that chapter. To some extent, journalism has been a subject for study *per se* and has an independent literature which might be thought of as warranting review perhaps as an initial source of data. However, the journalism literature tends to operate under a sociological mandate rather than a focus on the design of technological tools. It also tends to be based on large scale surveys offering limited theoretical richness, or alternatively is intuited rather than researched and can tend to be idealised. Consequently, it was considered that a preferable source of data about the process of journalistic

writing would be observation (Chapter 3) and interview (Chapter 4). Hence, journalism literature is not included in the review.

The review divides into four main sections. Section 2.2 looks at research in Information Science concerned with information-seeking and use. Since this has central relevance for this thesis it is the largest section of the review. This literature is extensive and so the review is selective. The section focuses on important contributions and on work that provides useful context for this thesis. Taking a broadly chronological approach, the section tracks research developments within the user-centred paradigm in Information Science paying particular attention to some key themes (which for the most part, run as threads throughout). These themes are:

- Information-seeking context, and in particular, task context;
- The drive for the development of theory;
- Cognition and the Cognitive Viewpoint;
- Information-seeking as an evolving process;
- Information-seeking uncertainty;

Section 2.3 then reviews Information Science research which has been more specifically focussed on journalists and journalism.

Meta-theoretical debates rage in HCI as they have done in Information Science and to some extent developments in both have been affected by changes which have run through the behavioural sciences more generally. In section 2.4 the themes of cognition and the context of behaviour are reintroduced and developed through a review of three approaches that have been influential in HCI which provide valuable perspective for the work in this thesis; these are: Situated Action, Distributed Cognition and Cognitive Systems Engineering.

Finally, the context for the information behaviour studied in this thesis is a kind of writing task, and research into how people write may be valuable for the interpretation of observations of journalists. The Cognitive Science literature has provided a number of models of writing, and, in section 2.5, two of the most influential models are described.

2.2 Information-seeking in Information Science: cognition, context and uncertainty

Over the past forty or so years, user-centred research into information-seeking in Information Science has attempted to place the information service user at the centre of investigations into information needs, information-seeking strategies and behaviours constituting information-seeking processes. The emergence of a user-centred approach is frequently contrasted with the older 'Cranfield Paradigm', or system-centred approach (Ellis, 1992) which instead focuses its attention on the design and evaluation of search engines.

Arguably, the Cranfield Paradigm has provided user-centred information-seeking research with something of a sense of itself by providing it with a sense of what it is not, and so it will be useful to briefly acknowledge the system-centred approach here. Empirical IR research began at the Cranfield Research Institute in 1957 where procedures were developed for comparing the performance of IR systems. These evaluation procedures were based on lab-tests using standardised recall[1] and precision[2] test metrics. The test procedures for system comparisons developed at Cranfield, and the recall and precisions metrics which they used were, and continue to be, very influential. However, users were not represented in the Cranfield tests save for some implicit assumptions about IR system use and relevance judgements. Some of these assumptions have since been brought into question within the user-centred paradigm and enter into the discussion below.

2.2.1 Paisley's (1968) review

The term Information Science had only been around for ten years when Paisley published his influential 1968 review of research into information-gathering and dissemination by scientists and technologists[3] appearing in the *Annual Review of Information Science and Technology* (ARIST) between late 1966 and the end of 1967.

---

[1] Recall is defined as the proportion of the documents relevant to a problem within a collection which are returned by a search engine on some query.
[2] Precision is defined as the proportion of the documents returned by a search engine on some query which are relevant to a problem.
[3] The concern with technologists and scientists reflects the early aims of Information Science which had been to further the communication of scientific research (Farradane, 1970, cited in Ingwersen and Pors, 1995).

Paisley classified papers according to a framework in which the scientist/technologist is seen at the centre of any number of psychosocial systems. For example, they included *the scientists within a culture*, *within a professional membership group* (*e.g.* psychologist), *within an invisible college* (*i.e.* a small group of scientists who know each other and share information directly), *within a work team* or *within his own head.* The scope of these research concerns is suggestive of a holistic and potentially rich approach to information behaviour research. However, for the most part research was not concerned with the complexities of how people operate and work within these systems, but with more limited issues such as the channels of information flow and associated user likes and dislikes. Paisley criticised what he called the 'shallow conceptualisation' of the user in these studies and argued that in many studies it was "hard to glimpse a real scientist or technologist at work, under constraints and pressures, creating products, drawing upon the elaborate communication network that connects him with sources of necessary knowledge" (Paisley, 1968, p.2). In particular he noted a failure to consider:

1. The full array of information sources available;
2. The uses to which information will be put;
3. The background, motivation, professional orientation and other individual characteristics of the user;
4. The social, political, economic and other systems that powerfully affect the user and his work;
5. The consequences of information use—*e.g.* productivity;

Paisley called for more eclectic data gathering methodologies, recommending, for example, the use of questionnaires, diaries and structured interviews. Importantly, Paisley also noted that theory and its development appeared to play a minimal role in research in the area at that time. He cautioned against this, arguing that findings that did not reference theory or contribute to its development were in danger of being 'scattered' and 'disorganised'.

2.2.2 Levels of Information need - Taylor (1968)

A theory which was mentioned only in passing by Paisley, but which has been highly cited since, and in some ways anticipated and contributed to the 'cognitive turn' as it occurred in Information Science (discussed in section 2.2.4

below) was Taylor's levels of information need. Taylor analysed questions and subsequent negotiations between library users and librarians during reference interview situations, and, on the basis of this analysis, defined four levels of information need reflecting the process of moving from the actual (but perhaps unrecognised) need for information to an expression of a need which could be presented to an information system. Taylor's levels were:

- **The visceral level** - a vague sense of something missing;
- **The conscious level** - a clear need for information but an inability to express precisely what is sought;
- **The formal level** - an ability to state what information is needed;
- **The compromised level** - where the expression of information need is modified to accommodate an available resource;

Central to Taylor's model was the idea that enquiry is a process with identifiable stages and that information service users often have unclear ideas of what it is that they are looking for. Taylor also emphasised the idea that, in order to resolve an information need using some information resource, the expression of that need must be compromised in some way. Ultimately, information-seeking is a practical business with an information seeker needing to exploit some available resource to meet their needs, and this inevitably imposes constraints on how a question or query can be expressed.

2.2.3 Dervin and Nilan's (1986) review

Eighteen years later little appeared to have changed. 1986 saw the publication of another much cited review of ARIST papers by Dervin and Nilan who reasserted many of Paisley's arguments for change. Dervin and Nilan focussed on what they termed "the conceptualisations that drive research" (Dervin & Nilan, 1986, p.1), by which they meant the meta-theoretical views or orientations that underpin research concepts and methodologies. In this respect, they identified a tension between empirical studies, which for the most part resembled an old-style approach, and critical essays, which called for (and attempted) "a reassessment of what information needs and uses research is about" (Dervin & Nilan, 1986, p.4). This reassessment advocated two key developments. The first, echoing Paisley, was a shift in focus to more holistic views of the situations that give rise to information needs and processes. The

second was a shift in focus to consider the cognitive processes at work during information-seeking.

Part of this reassessment involved locating the information system user's needs and information uses as a focus for research and the driving force for design. In part, a focus on the user and their needs was perceived as involving a move away from large surveys classifying users in terms of broad demographics, and a move away from information-seeking research that was limited to the point of user/system intersection considered in the systems terms, to a richer understanding of users and the social, situational contexts that underlie their information-seeking behaviour. At the same time, the view of users as passive information recipients was being replaced with a constructivist perspective in which internal cognition was to become a key concern.

From a methodological perspective, it was proposed that the previous study format, which had focussed on the use of quantitative analyses, should be supplemented with inductive, qualitative research based on smaller numbers of users.

Dervin and Nilan summarised the 'traditional' study as one that focused on,

> ... research questions that start with the system—the source of the packages of information that are to be transferred from system to user. Such a study looks at how much use people make of these systems. It asks what demographic and observable sociological dimensions of people's lives predict this use. It is concerned with whether people are aware of these systems and like or dislike them. Its asks many "what" questions—*e.g.*, what people use what systems, and what services do people use.
>
> **(Dervin & Nilan, 1986, p16)**

In contrast, the 'alternative' paradigm,

> ... focuses on understanding information use in particular situations and is concerned with what leads up to and what follows intersections with systems. It focuses on the user. It examines the system only as seen by the user. It asks many "how" questions—*e.g.*, how do people define needs in different situations, how do they present these needs to the system, and how do they make use of what systems offer.
>
> **(Dervin & Nilan, 1986, p16)**

Dervin and Nilan finish their paper with a review of three approaches which, in their view, adopted this 'alternative' paradigm. These were Taylor and

Macmullin's User-Values approach, Dervin's own Sense-making approach, and Belkin's Anomalous States of Knowledge (ASK) approach. The latter two of these three have been particularly influential will be discussed below.

2.2.4 The Cognitive Viewpoint - De Mey, Ingwersen, Belkin, Oddy

Paisley (and others) had called for greater theory development in Information Science as a means of constructing more coherent research efforts. Devin and Nilan's concerns included developing greater focus on the cognitive processes underlying information-seeking behaviour. Influenced by the growth of Cognitive Science and its focus on the representation of knowledge and associated processes, an important meta-theoretical perspective for Information Science research, which, it was claimed, could provide a universal basis for theoretical progress, came in the form of the 'Cognitive Viewpoint'.

Key proponents of this view have been De Mey, Ingwersen, Belkin and Oddy. In 1980, De Mey declared that the basis of the Cognitive Viewpoint in Information Science was that:

> ... any processing of information, whether perceptual or symbolic, is mediated by a system of categories or concepts which, for the information processing device, are its model of the world.
>
> **(De Mey, 1980)**

Central to this view is the imperative of paying attention to the various cognitive or 'knowledge' structures of people involved (historically or otherwise) in structuring and determining IR processes. In this are included authors, system designers and indexers, as well as information workers and users.

Figure 2.1 shows Ingwersen's model of factors influencing IR interaction (Ingwersen, 1992) which represents what for Ingwersen is the scope of the Cognitive Viewpoint. In it he includes implemented structures representing the designer's conceptions of how to process information objects in the system, system objects which include index representations of text or pictures as well as the text or pictures themselves, the intermediary (person or mechanism) mediating the user's access to the system, the user, and their social/organisation environment.

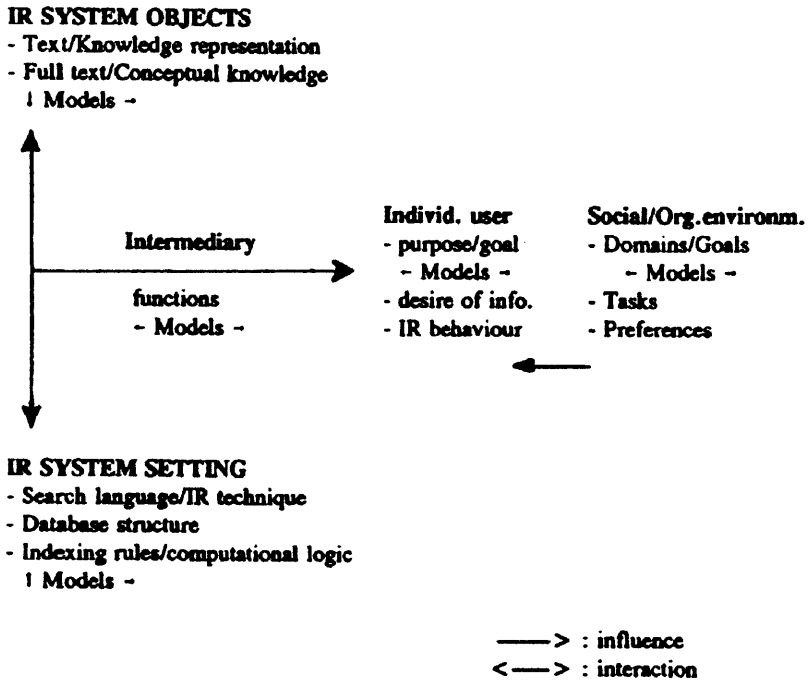


Figure 2.1 Ingwersen's (1992) model of the scope of the Cognitive Viewpoint

Ingwersen's vision of the Cognitive Viewpoint, then, is that it should be inclusive, integrating multiple elements and perspectives in the IR situation represented in terms of interacting cognitive structures.

Ingwersen associates the Cognitive Viewpoint with other cognitive based approaches such as those incorporated into Cognitive Science, but goes to some length to distance it from *cognitivism* with which he claims the Cognitive Viewpoint is frequently confused (Ingwersen, 1992). Cognitivism, he argues, corresponds with 'Strong AI' (Searle, 1984) and the view that computers can literally have thoughts and feelings. Ingwersen's position is that computer systems process information imbued with the cognitive structures of their designers, but that their operation occurs at a purely syntactic level.

Taylor's levels of need 1, 2 and 3 cognitively contextualise the observation that users frequently have difficulty in articulating their needs. In order to provide an account of information need in terms of inner cognitive structures, Belkin, Oddy and Brooks (1982a, 1982b) offered their ASK hypothesis. The ASK hypothesis states that:

> .... an information need arises from a recognized anomaly in the user's state of knowledge concerning some topic or situation and that, in general, the user is unable to specify precisely what is needed to resolve that anomaly.
>
> **(Belkin, Oddy & Brookes, 1982a, p.62)**

The idea of an *anomalous state of knowledge* stems from the idea of human interaction being mediated by knowledge structures, or what they termed "the conceptual state of knowledge" (Belkin, Oddy & Brookes, 1982a, p.65). An *anomaly* in their sense is the recognition of some inadequacy in a conceptual state of knowledge with respect to some aim. The problem of information need uncertainty is a central component of the ASK hypothesis and has gained considerable importance within the information-seeking literature. Given users' difficulty in articulating their needs, Belkin *et al.*. argued that supporting this is the fundamental issue in IR and consequently the place where system design should begin.

Adopting Wersig's (1979) idea of a 'problem situation' to mean the wider tasks or situation that brings someone to seek information (as an example of such a problem situation they describe the goal of performing an evaluation of the information services of professional institutes, where information services include publications, conferences, seminars, exhibitions, meetings, enquiry services *etc.*), Belkin *et al* argued that rather than users having to say what information they wanted, a system should allow them to describe their problem situation—an easier task for users in their view. They explored a format for representing documents and problem situations in a way that would permit matching between the two using networks based on term associations derived from documents and statements of problem situations. An evaluation of this idea based on users' and authors' views of the representational correspondence between some example networks and source problem situations and documents showed moderate success.

The idea of treating a representation of terms and their relationships in text as equivalent to a representation of cognitive structures, though, has come under some criticism under the argument that it is not obvious that the two are equivalent (Ellis, 1990). On a more general level, the Cognitive Viewpoint has also come under some criticism with one of the most consistent criticisms being that it fails to accommodate the holistic, situational mandate which had been set out by Dervin and Nilan and also by Paisley. For example, in a polemic

discourse analysis of Cognitive Viewpoint literature, Frohmann (1992) challenges the focus on mental structures or schemes arguing that:

> .... The ideology of interiors and its sharp opposition between inner and outer underwrites a related discursive strategy of radical individualism
>
> **(Frohmann, 1992, p.375)**

...and that:

> The erasure of the social thus becomes one of the Cognitive Viewpoint's most significant discursive achievements.
>
> **(Frohmann, 1992, p.376)**

In the social constructionist approach, information use is crucially seen as more contextually bound (Tuominen & Savolainen, 1997). Similar criticisms concerning the lack of account made of context by the Cognitive Viewpoint have been made by Ellis (1989a) and Lueg (2002).

These arguments turn on perceived limitations in the Cognitive Viewpoint in terms of the relative importance placed on accounts of 'inner' life in accounting for what people do, at the expense of 'outer' life—that context plays an essential role in determining human behaviour and that a cognitive scope fails to accommodate this. If it can't conceptualise the socio-cultural context then it simply remains "a theory of how individuals process information" (Talja, 1997, p.67). In response to such criticisms however, Ingwersen has argued that, as far as the Cognitive Viewpoint is concerned, "... one may state that no *theoretical* limitations exist for exploding the system of environmental variables—only operational and methodological limits" (Ingwersen, 1992, p.19, authors emphasis). In making this point Ingwersen draws attention to the right hand side of his scope model shown in figure 2.1 in which he identifies external factors including the social and organisational environments as amenable to cognitive analysis. For Ingwersen, the Cognitive Viewpoint not only provides a common theoretical perspective suitable for the study of user-cognition, but it also provides the possibility of contextual richness.

2.2.5 Sense-making theory - Dervin

Dervin's work on Sense-making has been classed by Ingwersen as a departure from the Cognitive Viewpoint towards a more communication based approach

(Ingwersen, 1992), although Dervin herself has described it as standing "between some traditional, frequently illusionary and restraining polarities" (Dervin, 1983, p.4). Sense-making theory, however, replaces the Cognitive Viewpoint's focus on mental representations with the concept of dialogue (Dervin *et al.*, 1992; Dervin, 1994), and it has been thought of as closely corresponding to the discourse analytic viewpoint (Talja, 1997).

With the aim of providing "a coherent set of concepts and methods" for information needs research (Dervin, 1983, p.3), Dervin places an emphasis on the situations in which users find themselves when information needs occur, including in this any situation where information is needed to achieve some end. Sense-making is concerned with the study of "how people construct sense of their worlds and, in particular, how they construct information needs and uses for information in the process of sense-making" (Dervin, 1983,m p.3). The approach is based on a SITUATIONS-GAPS-USES model of information need. That is, an information need is a kind of 'gap' in someone's understanding—a gap that stands between them in their current situation and information uses (information uses being goals that are achieved through a newly constructed sense of some aspect of the world). According to this view, information, rather than being a static external representation, is seen as the dynamic, subjective, internally constructed sense that is made of a message. Hence, information-seeking is a dynamic activity in which the user is constantly updating their own internal structures as new information is acquired and is assimilated into what is already known.

In terms of methods that have been used to understand the nature of these gaps, Sense-Making work has chiefly focused on the use of the Micro-Moment Time-Line Interview. In this technique, the subject is asked to detail a step-by-step, post-hoc account of what happened in a situation. Then, for each step, they are asked what question they had at that time, what things they needed to find out, learn, or make sense of, and, if an answer was obtained, the ways in which it helped or hindered their situation. The aim is to establish two or three dimensions of the SITUATIONS-GAPS-USES model. In-depth analyses are then carried out depending on the purposes of the study.

Dervin acknowledges a debt to, among others, communication theorists who have taken a situational, constructivist approach to studying communication as

behaviour, and most notably Carter and his ideas of communication as gap bridging behaviour. For Dervin, information-seeking and use are constructing activities which make successive modifications to one's internal picture of reality—thus she emphasises the subjectivity of interpretation. The view is not so much that people are passive recipients of information—that they are informed by it—but rather that people use information to construct new understandings with the aim of moving closer to achieving their goals.

Constructivism has become an increasingly important view in the Information Science literature on the cognition of information-seeking (see also for example Kuhlthau 1993, Ingwersen 1996), and it will be useful to digress slightly to consider its claims. Constructivism, as an approach, grew out of the developmental psychology of Piaget and Russian educational psychologist Vygotsky. Vygotsky's theories emphasise discovery and cooperative learning, and that children learn to solve problems by thinking their own way through them. Central to the constructivist view is the idea that learning is an active rather than a passive process in which the learner engages in building their own concepts to enable an understanding of the world around them. This process of assimilation implies that past learning experiences and the concepts that the learner has actively built in response to these experiences provide the framework through which new experiences are interpreted and integrated into the developing cognitive structures.

Related to this, Dervin's view embraces the notion of evolving information needs—as one gap is bridged and a new understanding is constructed, so another gap may be revealed and so on..... . Central to Sense-Making is the idea that behaviour can be predicted more successfully within a framework of changing situations as predictors.

Also, her conceptualisation of an information need (as a gap) is perhaps more intuitive than Belkin, Oddy and Brooks' anomaly. An anomaly is an exception to a trend or expectation and seems less obviously applicable to the notion of an information need than a gap. However, the SITUATIONS-GAPS-USES model appears to capture a notion of uncertainty less adequately. Here a criticism is levelled at Dervin's model on its own terms—as a model capturing the sense-maker's perspective. The problem is that the notion of a gap lying between situations and uses could be taken as implying that use is in some sense

predetermined and hence that a necessary bridging function is predetermined too; in reality this can be far from the case. There will undoubtedly be gaps and there will be uses, but, arguably, from the sense-maker's perspective (and in a Wittgensteinian sense) their constructs are the limit of their world. Hence, uses (and bridges) lie beyond the visible horizon, and, from the sense-maker's perspective, are far from determinate.

2.2.6 The Behavioural Approach - Ellis, Bates

As another alternative to the cognitive approach, Ellis has proposed what he calls a 'behavioural' approach to modelling information-seeking situations and ultimately of understanding how to design better IR systems (Ellis, 1989a). Rather than focussing on cognitive structures involved in IR interaction as the objects of research, the Behavioural Approach is presented as a more holistic perspective (Ellis & Haugan 1997) centring around the description of prototypical information-seeking behaviours of users in the context of a rich description of the work task context. Ellis' rationale was that,

> If researchers' information seeking patterns are broken down into their basic behavioural characteristics – and the retrieval system is provided with facilities that reflect those characteristics – then users should be able to recreate their own information seeking patterns while interacting with the systems.
>
> **(Ellis, 1989a, p.172)**

Ellis began by deriving a model of the information-seeking activities of social scientists based on semi-structured user-interviews designed to enable detailed accounts of the perceptions of this group "from their point of view, and as a whole" (Ellis, 1989a, p.172). He adopted Glaser and Strauss' (1967) Grounded Theory approach in order to abstract behaviour patterns and guide sampling. The resulting model consisted of six 'characteristics' of the social scientists' information-seeking behaviour. These were:

- **Starting:** Depending on experience on a topic, alerting the searcher to principal ideas or key studies to provide an overview or introduction to an area;

- **Chaining:** Exploiting citations in academic articles, chaining is the practice of tracing from a given article to articles that it cites (backwards chaining) or tracing from a given article to articles that cite it (forward chaining);
- **Browsing:** Semi-directed or semi-structured searching in an area of potential interest based on the collocation of like material;
- **Differentiating:** Exploiting known differences between sources as filters on the nature and quality of the material examined;
- **Monitoring:** Maintaining awareness of developments in a field through the monitoring of particular sources.
- **Extracting:** Systematically working though a particular source to locate material of interest.

Ellis pointed out that these characteristics were abstractions based on particular practices, and that any particular instance might exhibit more than one characteristic.

Despite the model being initially grounded in the activities of a specific user population, and despite the fact that it managed to embrace contextual richness present in that situation, at the model's highest level (outlined above) it has also shown itself to be generalisable to users in other work domains with very little modification. These domains include physical scientists (Ellis, Cox & Hall, 1993) and industrial research scientists and engineers (Ellis & Haugan, 1997).

There is, however, a potential limitation in studying current practice to inform future design. At any given time, goal directed behaviour depends upon the resources available, and specifically, what these make possible and easy. Consequently, new possibilities would undoubtedly reshape behaviour and reveal a different set of needs. This is a general problem for interaction design and it means that basing new systems on existing behaviour is likely to result in evolutionary rather than revolutionary solutions. However, two factors contributed to Ellis' approach providing useful requirements data for information system design. First, the 'observational' scope extended beyond the use of computerised information systems to information-seeking in general, and hence it was possible to consider how technological systems might be designed to support more activities. Second, where users described activities supported by existing IR systems, these were not necessarily *well* supported. Hence in this situation design improvements can be made.

2.2.7 Berrypicking – Bates (1989)

An important theory falling within the Behavioural Approach (Ellis, 1989b), which elaborates the idea of need evolution in information-seeking, is Bates' (1989) Berrypicking model. Bates referenced her work against perceived shortcomings of the Cranfield Paradigm and argued that the Berrypicking model was a more accurate representation of real information-seeking behaviour. Bates contended that the IR literature had been preoccupied with subject searches conducted over bibliographic abstracts and indexes, and that this formed only one of a number of strategies that were widely used. Significantly, she characterised the traditional system-centred model as assuming, or at least being limited to, static, isolated searches in which a single, unchanging need determines a single query to be matched against a document set. Where the traditional model *did* accommodate the idea of query evolution, such as in Salton's relevance feedback approach (Salton, 1968), it was nevertheless assumed that the underlying need was unchanging.

Against this view, Bates argued that users move through a variety of strategies using different information sources, and, crucially, as information is found, so this provides new ideas and directions to follow; hence information needs evolve. Central to the model was the idea that rather than one single results set being useful to the user, users select from the information that they encounter throughout the process—hence the use of the 'Berrypicking' metaphor. Figure 2.2 (reproduced from Bates, 1989) represents the evolving Berrypicking search as a path. Within the figure, changes in direction of the path represent changes in the user's need corresponding with following different leads and shifts in thinking. Occasional arrows to documents indicate documents and information being produced from the search at points along the way.

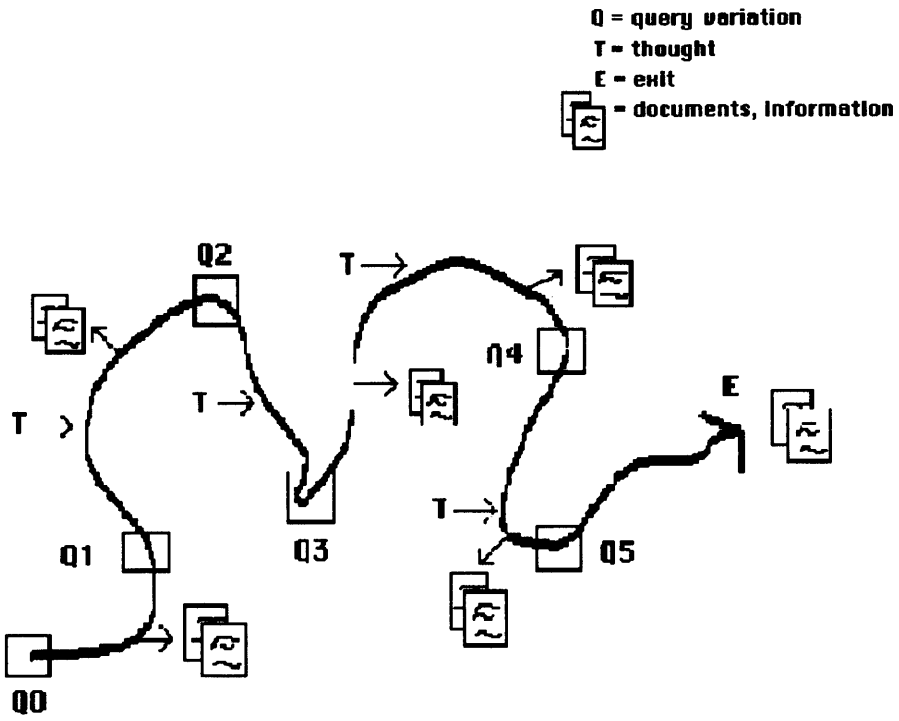


Figure 2.2 Bates' (1989) Berrypicking Model

2.2.8 The Information Search Process (ISP) model - Kuhlthau

Like Dervin and Nilan (1986), Kuhlthau has also emphasised the importance of understanding information-seeking as located within a task context (Kuhlthau, 1993; Kuhlthau & Tama, 2001). However, Kuhlthau and Tama have recently noted that, for the most part...

> ...information systems and services have been designed to support information seeking and gathering without consideration for accommodating the ultimate need of applying information to accomplish work tasks.
>
> **(Kuhlthau & Tama, 2001, p.26)**

Belkin, and to some extent Taylor, had emphasised the idea that information seekers are often uncertain about their information needs and Bates had observed information-seeking as an evolving process. In Kuhlthau's work both a link is made between uncertainty and evolution through her ISP model (Kuhlthau, 1993).

Kuhlthau focused on user cognition, behaviour and, importantly for her, the associated affect of information seekers during the search processes conducted in the context of wider tasks. She described a model of information-seeking based on constructivist theories of learning by Dewey, Kelly and Bruner, and a series of studies of information-seeking (primarily of school and university students) from their own perspective (Kuhlthau, 1988a; 1988b; 1988c; 1989).

The ISP model (shown in figure 2.3), which locates information-seeking within the context of wider task goals, describes the information search process as a series of stages, with each stage being a prerequisite for the next.

Initiation is the point at which a person first becomes aware of a lack of knowledge.

Selection involves choosing a topic for an assignment. This is decided by balancing constraints such as personal interest, assignment requirements, information available and time allotted. During this phase background information is sought about the areas under consideration.

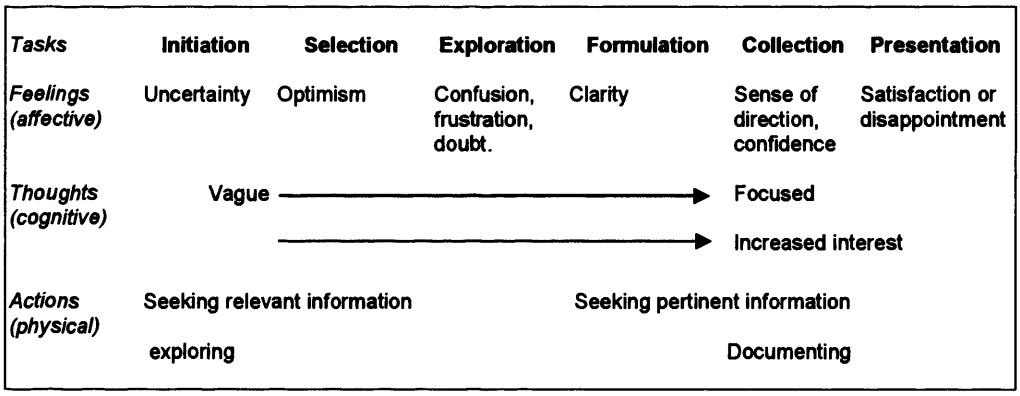


Figure 2.3 Kuhlthau's (1993) ISP model

Exploration is where a further focusing decision is made, which is to select a topic within the chosen area. The aim of information-seeking at this point is to become informed enough to form a point of view. Importantly, information needs are uncertain at this point and so expressions of needs are imprecise. Interactions with information services are consequently problematic.

Formulation is where a focus emerges.

Collection involves gathering information about the focused subject. Users have a clearer sense of direction and can specify needs more precisely. Interactions with information services are at their most effective.

Presentation is where the search is completed.

Central to the ISP model, then, is the idea that information needs begin as vague and unclear, and interaction with information services is difficult. Later, as the user becomes more focused, so their searching becomes more focused and precise. The picture of evolution presented is one in which increased knowledge and understanding leads to more specific plans for the higher-level task and a consequent narrowing of scope in information needs and greater success with information providing services .

Kuhlthau's model adds to the ideas of Belkin and Taylor by identifying circumstances under which information needs tend to be uncertain and difficult to articulate. It also states that uncertainty at the level of need articulation is a function of uncertainty at the level of the underlying task. Kuhlthau put forward a challenge to information-seeking system designers remarking that library and information systems based on the bibliographic (Cranfield) paradigm are best suited to document collection where the user has a clear idea of their needs (*i.e.* the later stages) and that they ignore the "holistic experience of information seeking" (Kuhlthau, 1993, p.352). In particular they do not support the user during the active constructive process of exploring information and formulating an understanding.

2.2.9 Information behaviour - Wilson

Wilson has been a strong advocate of research into information-seeking being located within a rich understanding of the context of use and has recently identified the lack of this kind of research.

> ...[in 1981] information use had received little attention and, within Information Science, that statement is still relatively true today.
>
> **(Wilson, 1999, p.251)**

In common with Dervin and Nilan (1986), Wilson also emphasises the value of qualitative data gathering techniques arguing that such methodologies tend to make a greater contribution to the development of theory than positivist approaches (Wilson, 1999).

In arguing for a widening of the scope of information-seeking research, Wilson has defined the term *Information Behaviour* as referring to,

> ... the totality of human behaviour in relation to sources and channels of information, including both active and passive information seeking, and information use.
>
> **(Wilson, 2000, p.49)**

This broadening of the unit of analysis beyond information-seeking to information-seeking and its use accords with the Dervin and Nilan idea of extending the research perspective beyond the point of system interaction. Importantly, it places an emphasis on information-seeking motivation, *i.e.* the user's wider task.

In an attempt to extend and integrate existing information-seeking models into a model of information behaviour revolving around a task context, Wilson presented a model based on the notion of wider tasks as problem solving (reproduced in figure 2.4). This model was subsequently adopted and evaluated as a framework for the Uncertainty Project (Wilson *et al.*, 2002).

Wilson's model decomposes problem solving into four consecutive process stages: problem identification (where the person asks, 'What kind of problem do I have?), problem definition ('Exactly what is the nature of my problem'), problem resolution ('How do I find the answer to my problem'), and potentially, solution statement ('This is the answer to the problem').

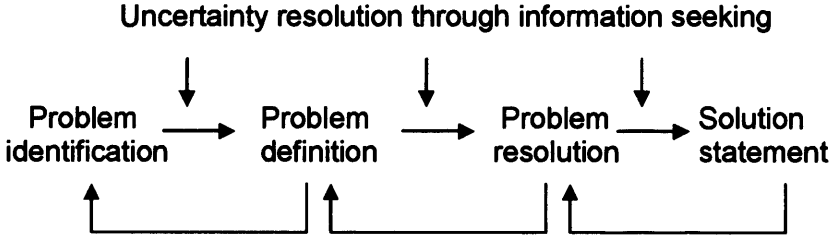


Figure 2.4. Wilson's (1999) problem solving model

According to the model, information-seeking may be conducted at each stage of problem solving and may contribute to the resolution of uncertainty and transition to a subsequent stage. Wilson suggests that the model can subsume other well known models of information-seeking, such as the ASK model (Belkin, Oddy & Brooks, 1982a) and the Information Search Process (ISP) model (Kuhlthau, 1993). He also associated Ellis' search characteristics with lower-level process elements at each stage. The possibility of new information contributing to an increase in uncertainty and returning the problem solver to a previous problem stage, is represented by feedback arrows.

2.2.10 Task complexity - Byström and Järvelin

Another approach to conceptualising information-seeking task context was provided by Byström and Järvelin (1995) as part of a study into the relationships between task complexity and factors such as the kinds of information sought, and the information channels and sources used. Byström and Järvelin developed a framework for classifying tasks according to levels of complexity based on characterisations that had been used in psychology, organisational studies, and other information-seeking studies. The characteristics of complexity that they considered included: *a priori* determinability, receptivity, analysability, the number of alternative paths of task performance, outcome novelty, number of goals and conflicting dependencies among them, uncertainties between performance goals, number of inputs, cognitive and skill requirements, and time-varying conditions of task performance. In considering these, Byström and Järvelin judged them as belonging to one of two main groups—characteristics related to *a priori* determinability and characteristics related to the extent of the tasks. Of these, the dimension that they choose to operationalise as a measure of complexity was *a priori* determinability. This feature, as an identifier of wider task complexity, has been most widely used in information-seeking research (Vakkari 1998, Vakkari 1999).

Byström and Järvelin defined lack of *a priori* determinability as uncertainty concerning task outcomes, process and information requirements. This is essentially an adaptation of a concept from Van de Ven and Ferry (1980) (also adopted by Tiamyu (1992)), which states that task complexity relates to uncertainty associated with the inputs, procedures and outcomes of a task. The adaptation is that Byström and Järvelin substituted 'information requirements' for

'inputs'. So complexity becomes a function of the *a priori* determinability of information requirements, process, and outcome.

Using the concept of *a priori* determinability to characterise complexity, Byström and Järvelin classified task types as ranging from *automatic information processing tasks*, which are *a priori* completely determinable (and could, in principle be automated), to those that they refer to as *genuine decision tasks*, which are:

> ...unexpected, new and unstructured. Thus neither the result, the process, nor the information requirements can be characterized in advance. The first concern is task structuring.
>
> **(Byström & Järvelin 1995, p.7)**

Finally, they make the point that the level of complexity, or *a priori* determinability, is relative to the point of view of the user. Task complexity is not fixed for a given task, but concerns the relationship between the task and the user's knowledge and expertise. This intuitive point has been made by a number of researchers in the area of information-seeking and is reinforced by Ng (2002) in a study exploring the extent to which people plan information-seeking interactions as opposed to adopting more situated responses. Ng found that for subjects with higher system knowledge or higher subject matter knowledge there was less observed deviation from preformed plans. Thus assignment uncertainty is a function of the relationship between the task and the user.

Byström and Järvelin's study looked at the information-seeking behaviour of civil servants working in a local authority office in Finland. Using their complexity based classification, they found that as complexity increased, so the complexity of the information needed increased, the proportion of domain information (general facts) increased, as did the proportion of problem solving information (methods), the success of information-seeking decreased, selected channels became more external and the number of information sources increased.

2.2.11 Developing the theme of focus formulation - Tang and Solomon, Yang, Vakkari

Following Kuhlthau's work, there has been growing acceptance that formulation, in respect of a wider task context, represents a particularly significant point for information behaviour. Obtaining a focus on their wider task is understood as a

turning point for the user. Prior to it, uncertainty wavers, but generally increases in intensity; after, uncertainty subsides and confidence increases.

Vakkari, who adopted Kuhlthau's ISP model as a framework for his own research, interprets formulation as follows:

> The finding of a focus is crucial in the search process. The focus is comparable to a hypothesis for accomplishing the task. Prior to formulation thoughts are general, fragmentary and vague, and actions involve seeking background information... After a focus has been constructed, the search for information becomes more directed. Thoughts about the task become clearer and more structured.
>
> **(Vakkari, 2001, p.46)**

Byström and Järvelin (1995) articulated formulation as creating a solution space and determining information requirements. As Vakkari (1999) argued, after formulation, the information seeker has a problem that might be solved, and knows more clearly what information is relevant.

Several studies have explored the progressive reduction in uncertainty as observed through two phenomena of IR interaction—query articulation and making relevance judgments. Tang and Solomon (1998) presented a case study of a single graduate student searching for documents in preparation for writing a term paper. In two observation sessions she was asked to mark retrieved records and journal articles for relevance. Tang and Solomon observed that during the first session, and after having evaluated an initial 16 items of a bibliographic results set, she asked if she could go back and re-evaluate the previously evaluated records. During her re-evaluation it became clear that her relevance criteria had become more focused. She commented that she had developed a better idea of what she was looking for and that her original selections were too general. At the second observation session, conducted after the subject had read her selected papers, she commented that a new topic had emerged during her reading. Content analysis showed that the new topic reflected the original retrieved set more closely than her previously chosen topic. It seemed that the subject had modified her topic in accordance with the opportunities presented by the available documents.

Yang (1997) reported a study of undergraduate students performing information-seeking using a hypermedia database in order to write a class assignment. The system used (Perseus) permitted users to retain items for later use, and hence provided a valuable opportunity for observing relevance judgments. Yang noted that, on occasion, subjects were uncertain about the value of some information and would sometimes defer judgment. Subjects made comments such as, *"I'll come back to this later"*, or *"I'll have to think about that"*. One subject said,

> Eventually..., this is the reconstruction of the Frieze..., is something I'm going to use... I'm pretty sure... I don't know at this point, so I think I'm going to think about it a little more... mmm...
>
> **(Yang, 1997, p.84)**

These studies demonstrate the development of a focus as an important factor impacting on information-seeking by demonstrating increased confidence and the ability to make categorical relevance judgments as a clearer idea of the goals of a wider task evolve. They support the idea of focus corresponding to a broad plan or goal and are consistent with Vakkari's notion of a hypothesis for accomplishing the wider task and Byström and Järvelin's idea of narrowing the solution space. As Yang argued in relation to one subject's exploratory information-seeking at an early stage of the task,

> It seemed that he [Eric] had no specific goal or coordinated plan in mind. He appeared to be exploring the database in hopes of hitting on something that might trigger an insight or idea.
>
> **(Yang, 1997, p.83)**

Similar results have been reported by Spink *et al.* (2002).

In a larger, longitudinal study, Vakkari *et al.* (Vakkari, 2000a; Vakkari & Hakala, 2000; Vakkari, 2000b; Vakkari & Pennanen, 2001; summarised in Vakkari, 2001) observed this same effect, and also demonstrated that obtaining a focus for a wider task facilitates greater query specificity. Adopting and refining Kuhlthau's ISP model, Vakkari *et al.* report the evolution of search tactics, search terms, relevance judgments and sources by a group of students writing a research proposal for their masters theses. Vakkari *et al.* chose to condense Kuhlthau's original six stages into three: prefocus, focus and postfocus. Various

data were captured over three search sessions evenly distributed across a four-month period. This study showed that throughout the task the students' problem stages (according to the three-stage model) could be systematically related to the use of increasingly more specific search terms and discriminating relevance judgments. In the later stages, broader terms were dropped and the students adopted an increasingly large and more specific vocabulary.

2.3 Information-seeking in journalism

Motivated by disparaging reports of the online searching skills of end users Nicholas (1996a) made a comparison between the search activity of search intermediaries and their respective end-users at the Guardian and also at the House of Commons. The study used interviews, questionnaires and transactional log analysis to compare factors including: the breadth of command vocabulary used, the number of search terms, the range of query fields used, the extent of browsing, the numbers of query reformulations, speed of search, the numbers of sources used, and search success.

Focusing on the results that Nicholas obtained from the journalists, he found that they had a more limited search vocabulary and used fewer search terms compared with their respective intermediaries. They also used a more limited range of query fields, focusing mostly on subject searches, which Nicholas noted, typically used names as query terms. Browsing was extensive among the journalists. 16% of their searches involved the user browsing more than 100 records. However, browsing was equally extensive among the intermediaries studied. Nicholas also observed, that the journalists performed fewer search reformulations than the intermediaries.

Many of these measures suggest that the journalists studied conform to the negative stereotype that Nicholas set out to consider. However, he also found that the journalists were not slow searchers and tended to use multiple sources. Perhaps most notably, as a measure of success, Nicholas found that, search satisfaction among the journalists was high (although less so for women journalists). On another success measure—the number of zero hit searches and number or errors—Nicholas found that journalists were only marginally less successful than their intermediary counterparts (no statistical comparisons were

performed). Also, although the librarians made fewer system command errors, they tended to make more spelling errors.

In an interview based case study, Nicholas and Martin (1997) elicited information needs data from journalists working at *The Independent*, *The Sunday Times* and *The Guardian*. The study was structured in terms of a framework initially proposed by Line (1969 & 1974), and developed further by Nicholas (1996). The framework distinguished ten characteristics of information needs and four kinds of obstacles that can stand in the way of meeting them. These were:

Characteristics:

- Subject: Subject matter of the required information;
- Function/purpose: The wider motivation for seeking the information;
- Nature: Whether the need is for conceptual, theoretical, historical, descriptive, statistical or methodological;
- Intellectual level: The extent of knowledge required to understand the information;
- Viewpoint: The sympathies of the required information;
- Quality/authority: The extent to which the information can be relied upon.
- Currency/date range: The time period for which the information is relevant;
- Speed of Delivery: How quickly the information is required;
- Place of origin: Where the information originated;
- Processing/packaging: How the information is presented;

Potential obstacles:

- Time: Work time constraints
- Access: Availability of the information.
- Information overload: Having more information than can practically be managed.
- Training: The level of expertise in exploiting information resources;

Since the goal of the current thesis is to understand information-seeking and information management in the context of a wider task where, at one level, the wider task is journalistic writing, the results obtained by Nicholas and Martin under the characteristic heading *function/purpose* are particularly relevant. Nicholas and Martin decomposed this aspect of information needs into five kinds of purpose:

- **Fact-checking:** This might be wanting to know the names of companies involved in a merger, the date an article was published, or checking the wording of a quotation.
- **Current-awareness:** Generally keeping in touch with latest developments.
- **Researching:** Using documented material (in conjunction with current events) as the basis for a story.
- **To obtain context:** Finding background information with which to contextualise current events in a report.
- **Stimulus:** Serendipitous information finds which trigger a story idea.

(Arguably, this last category is not so much a purpose for information-seeking so much as an unplanned event.)

Nicholas and Martin concluded by emphasising the extent of the journalists' information needs, their requirement for authoritative and current sources, and the speed with which information is needed. The major obstacle (or challenge) that they experience is lack of time.

There are some identifiable similarities between the Nicholas and Martin functions and Ellis' behaviour characteristics. (In chapter 4, comparisons are made between behaviours identified in this thesis and the models of Ellis and of Nicholas and Martin). Probably the most striking similarity lies between Nicholas and Martin's *current-awareness*—the goal of keeping up-to-date with developments, and Ellis' *Monitoring*—maintaining awareness of developments in a field through the monitoring of particular sources.

2.4 Context and the HCI agenda

According to the ACM SIGCHI Curricula for Human-Computer Interaction...

> Human-computer interaction is a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them.
>
> **(Hewett *et al.*, 2002, p.5)**

Central to HCI's aims are the provision of methods and theories to support the design of more effective and satisfying interaction. To these ends, *major phenomena* can be understood as including the ways that people think about and perform activities with different kinds of devices. Consequently, developing an understanding of information behaviour during information-seeking and writing tasks in order to inform information system design can be understood as falling within the scope of HCI.

In HCI, particularly influential approaches for understanding how people perform tasks include hierarchical cognitive-task analysis techniques such as GOMS (Card, Moran & Newell, 1983) and TAG (Payne & Green, 1986). These essentially analyse tasks as action sequences by systematically decomposing them into lower-level sub-tasks. These approaches are particularly well-suited for analysing low-level, localised and well-defined behaviour sequences using specified devices under stable and predictable conditions. However, where goals and the available means for achieving them are uncertain and conditions are unstable, as in writing tasks, they are unsuitable, except perhaps for the analysis of low-level task components.

HCI has seen the recent emergence of more holistic approaches to modelling behaviour. These approaches, defined at least at a meta-theoretical level, are typically less formal and avoid assumptions of predictability in favour of the idea of discretionary problem-solving shaped by the contingencies of dynamic and unpredictable contexts. The thrust has been, on the one hand, to regard problem-solving as *ad hoc* adaptation to a dynamic situation, and on the other hand, to consider the structure of the environment as integral to how we think through problems. Hence, similar to Information Science, in HCI there is now often a greater emphasis on the need to account adequately for the role of context in determining task behaviour.

In this section, three context oriented meta-theoretical approaches are reviewed that have, to a greater or lesser degree, been influential in HCI and which will provide valuable perspectives from which to consider the aims and outcomes of this thesis. These are: Situated Action, Distributed Cognition and Cognitive Systems Engineering.

2.4.1 Situated Action - Suchman

In her book, *Plans and Situated Actions* (1987), Suchman criticised a view of action prominent in Cognitive Science which she referred to as the "planning view" (Suchman, 1987, p.27). Two of Suchman's principal arguments for the importance of accounts of behaviour being situated within accounts of context are summarised here and referred to as *the vague plan argument*, and *the indexicality argument*.

The vague plan argument - According to Suchman, in the planning view, "plans are prerequisite to and prescribe action, at every level of detail" (Suchman, 1987, p27). Consequently, any description of intention is reducible to a sequence of bodily movements. In the planning view, a plan is seen as the principal determinant of action in much the same way that a program controls the actions of a computer. Suchman's objection was that this overestimates the role of plans. She argued that plans are inherently vague, and that this vagueness not only makes plans tractable, but is necessary in order that intentions can be played out against the contingencies of an unpredictable environment. Suchman developed her view under the title of 'Situated Action'. Central to it is the idea that "every course of action depends in essential ways upon its material and social circumstances" (Suchman, 1987, p.50). Suchman argued that whilst plans are a resource for action they do not in any strong sense determine its course. To illustrate this she gives an example of preparing to run rapids in a canoe in which the canoeist might initially plan a broad trajectory, but, once the run begins, responding to the contingencies of the currents requires that they "abandon the plan and fall back on whatever embodied skills are available..."(Suchman, 1987, p52).

She takes the point further by arguing that:

> It is frequently only on acting in a present situation that its possibilities become clear, and we often do not know ahead

> of time, or at least not with any specificity, what future state we desire to bring about.
>
> **Suchman (1987, p52)**

Hence, plans not only lack implementation detail—the actions through which goals are achieved—there is also an inherent vagueness about the goals towards which our actions move us. The canoe example is, in part at least, an analogy and Suchman characterises the 'environment' of action as a series of social and physical situations that we walk into and to which we respond (and, in the case of social situations, also create), and the enumeration of these situations constitutes an account of situated human action.

The indexicality argument - Suchman was particularly interested in linguistic communication construed as a class of action and she starts the indexicality argument by using the indexicality of language as an example. The indexicality of language refers to the idea that the interpretation of an expression often depends upon the context of its utterance. For example, the use of pronouns, tense and place adverbs such as 'here' and 'now' mean that the meaning of an utterance can only be determined on the basis of the situation within which it was made. Suchman then extends the notion of the indexicality of language to all utterances and then (taking an utterance as a class of action), extends the argument further to say that understanding the significance of any action rests upon an account of the context in which it was performed.

> ... every instance of meaningful action must be accounted for separately, with respect to specific, local, contingent determinants of significance. The recommendation for social studies [and Cognitive Science] ... is that instead of looking for structure that is invariant across situations, we look for the processes whereby particular, uniquely constituted circumstances are systematically interpreted so as to render meaning shared and action accountably rational.
>
> **Suchman (1987, p67)**

Although Suchman was concerned with human-machine dialogue, the vague plan argument, with its notions of ill-formed plans and uncertain objectives, is particularly well-suited to human information behaviour and resonates very well with current views in Information Science about information behaviour in the context of complex tasks. Suchman's ideas are compatible with the notions of need uncertainty expressed by Taylor and Belkin *et al.*, and accord strongly with

the idea of need uncertainty in relation to task uncertainty as expressed by Kuhlthau and further explored by Vakkari and others. As Bystrom and Jarvelin pointed out, complex information tasks have associated *a priori* uncertainty and this is at the heart the vague plan (and goal) argument. The idea of engagement with information bringing about formulation and the greater clarification of task objectives is a prime example of Situated Action and the idea of the possibilities of a situation (and consequently desired future states) becoming clearer through acting. Information-seeking is almost by definition an opportunistic activity with unpredictable outcomes contingent on what a situation throws up. The unpredictable contingencies of currents in the canoeist example serve well as an analogy.

The indexicality argument is an argument about making interpretations and finding meaning in action, and is apparently orthogonal to the argument for the inherent vagueness of plans and goals. The indexicality argument argues that what an action *means* depends upon its context, and so to understand it properly we must provide an account of that context. Indeed, a hand raised to a friend at a station platform can mean one thing, but to an auctioneer it can mean something quite different, and to a passing taxi driver something else again.

The indexicality argument is more radically holistic than the vague plans and goals argument. It apparently not only argues for the embedding of accounts of behaviour within accounts of their context, but precludes the possibility of any account which abstracts behaviour from its context in any way. Hence, to subscribe to the *indexicality argument* is to deny abstraction, and, since generalisation rests upon abstraction, it is to deny generalisation too. This move is problematic for research, which relies upon generalising findings from one situation to another similar but inevitably different situation. But the argument also fails, and this can be demonstrated through an implied *reductio ad absurdum*, since the contexts within which phenomena naturally occur are effectively infinite, and yet finite accounts are all that are possible. There is no limit to what could be said about the context within which a phenomenon occurs but limits must always be imposed for any description to be given at all, and this must be true even of a situated account of human action.

The question, then, becomes one of where to impose the limits. In describing the context of human information behaviour we might, for example, describe the

kind of chair that the user is sitting on, but this would most likely be irrelevant detail since it would not (I am assuming) affect what they do or contribute to its interpretation. And so the question of where to draw the line on contextual description of behaviour hinges on a judgment made by the analyst about what factors affect the behaviour or contribute to an understanding of what behaviour it is. And choosing to dismiss some details as irrelevant, in essence, is abstraction which in turn enables generalisation. In choosing the factors to dismiss as irrelevant, of course, there is always a danger of dismissing or failing to see something important, and to the extent that this is the case, there will be limits on the explanatory power of the account and a danger of over-generalisation.

The descriptive challenge is to include all factors that are relevant and to exclude all that are not; and where the factors are diverse and many, a holistic, contextual approach is most appropriate. The holistic, anti-reductionist turn arises from the recognition that, in some situations at least, phenomena are typically embedded within many, complex and integrated factors such that these factors can radically influence the phenomena and so cannot be ignored. Consequently, explanation can only be made with reference to many issues. Conversely, a reductionist approach is more appropriate when one or two factors can be isolated as determining what is to be explained. The question hinges on the kind of phenomena being researched. In the case of information-seeking, and information behaviour in general, it seems impossible to divorce these from aspects of their natural context (such as the kind of work that motivates them) without losing a significant capacity to understand and explain them.

2.4.2 Distributed Cognition - Hutchins

The second theoretical approach reviewed here, which emphasises the role of context in structuring task behaviour, and has been influential in HCI, is *Distributed Cognition*. Distributed cognition is most closely associated with the work of Hutchins. The description given here is based on Hollan, Hutchins and Kirsh (2000).

Distributed cognition is presented as distinguishable from other approaches through its commitment to the idea that cognition extends across the boundary between inner and outer. It is argued that the traditional view of cognition is of

the manipulation of symbols in the head, and hence that the boundary of cognitive processes does not extend beyond the individual. In contrast, Distributed Cognition 'softens' this inner/outer division by delimiting cognitive processes in terms of *functional* rather than physical boundaries. Accordingly, the analytic scope of elements that participate in cognitive processes is expanded to optionally include groups of people and the physical world (including artefacts). Cognitive processes, it is argued, can be distributed across members of a social group, they can involve the coordination between internal and external (material or environmental) structure, and they can be distributed through time such that products of earlier events can transform the nature of later events.

From a Distributed Cognition perspective, cognition can involve the coordination between internal resources, such as memory, attention and executive function, and external resources—the objects, artefacts and at-hand materials which surround us. Rather than being viewed as simple stimuli for a disembodied cognitive system, work materials become elements of the system itself. In action people exploit and create structure in their environment, for example, in order to off-load cognitive effort. By extending the boundary of cognition beyond the individual, the individual is seen as an element in a complex cultural environment.

Distributed cognition, then, is a holistic perspective. Crucially, though, it is also cognitive. The idea of broadening the scope of a cognitive perspective resembles Ingwersen's ideas about the scope of the Cognitive Viewpoint in IR (discussed in section 2.2.4). Important differences between these two positions, however, can be understood by the distinct ways in which they deal with cognition with respect to external artefacts and cognition with respect to social systems. On the first point, Ingwersen regards artefacts such as documents and indexing structures as being, by design, imbued with cognitive properties. Perhaps we can understand by this as saying that these things are representational, or, as Belkin assumed for documents, they reflect the intentions and cognitive structures of their designers. Distributed cognition, however, is concerned with the extent to which the structural properties of environmental artefacts can be appropriated, harnessed and coordinated to support and, ultimately, be part of a cognitive process. Hence, the view is not necessarily one of artefacts being intrinsically cognitive or representational, but

of people using them as such in a more incidental, opportunistic or *ad hoc* fashion.

On the second point, Ingwersen regards social systems as relevant for IR research to the extent that they influence and modify the cognitive structures of people up to the point where IR interaction begins. In Distributed Cognition the importance of social issues lies in the extent to which it is useful to consider groups of people performing collaborative problem solving as one single cognitive system. For Ingwersen and the Cognitive Viewpoint, the social affects the cognitive (and presumably vice versa); in Distributed Cognition the social *is* cognitive.

2.4.3 Cognitive Systems Engineering - Rasmussen *et al.*

At the heart of Situated Action and Distributed cognition is the view that human activity is shaped and conditioned by the social and physical situation in which it occurs. Situated action emphasises an inherent vagueness of goals and plans and the ad-hoc adaptation to situations as they unfold. Distributed cognition emphasises the (potentially *ad hoc*) exploitation of situational structures to support and even be a part of problem solving. Hence, to create accounts of what people do, how they do what they do, and why they do what they do, we must know about the situations in which they do it. According to Suchman's indexicality argument, however, no contextual factor is given prominence over any other, and this lack of qualification brings with it the possibility for descriptions of irrelevant detail (as well as the need for infinite detail). The Cognitive Systems Engineering (CSE) (Rasmussen, Pejtersen & Goodstein, 1994) approach is a holistic framework for modelling work systems to inform technological system design, and it is introduced here since it provides demarcation of contextual factors relevant to these goals within the scope of complex work situations.

The aim of CSE is to model complex socio-technical work in order to predict how people operating within the system will behave in response to engineered changes. The goal is expressed as that of taking sociotechnical work systems apart in order to put them back together perhaps differently—to ask, what could be done differently and better? The aim is a top-down, step-wise narrowing of the degrees of freedom of prediction. A key premise of CSE is that in complex, dynamic work systems there are no predetermined procedures for doing work,

but rather that tasks are achieved through the discretionary exploitation and adaptation of available resources. Actors are given tasks and, given a range of resources, must choose how to perform them; and hence there is flexibility and adaptation on the part of the actor. As a consequence of this, prediction is described as problematic with action relying upon strategic flexibility in the face of the presenting situation (*i.e.* action is situated). CSE proposes that in order to address this problem, the analyst should systematically build a picture of a person's 'action alternatives' by modelling the boundaries of available choices.

The space of action alternatives is described as bounded (and consequently determined) by a set of active constraints and a set of available resources. In other words, constraints and resources set the boundary conditions within which work can take place and so shape peoples' choices and structure behaviour. Collectively, constraints and resources are referred to as *behaviour shaping constraints.* Behaviour shaping constraints, whether these be organisational, departmental or personal values, or resource possibilities and limitations, serve to determine the boundaries of available choices and so provide the delimitation of relevant from irrelevant aspects of context. CSE, then, demarcates factors which provide the focus for analysis, and yet it is holistic in the sense that, within the scope of behaviour shaping constraints, it extends to any issue whether it be cognitive, cultural or part of the material circumstances.

2.5 Writing process and writing technology

The task context for the information behaviour considered in thesis is news story authoring, and more generally, writing. This section considers literature related to the process of writing from the writer's perspective. Two influential models are discussed: Hayes and Flower (1980) and Sharples (1996). Also considered in this section is research which has explored the design and effects of different writing technologies.

2.5.1 A cognitive model of writing - Flower & Hayes

Flower and Hayes describe a cognitive model of expository writing (shown in figure 2.5) evaluated by the analysis of talk-aloud protocols elicited from writers. The aim of the model was to identify the cognitive sub-processes of writing and also to account for individual style differences. The model divides the writer's

world into *task environment*, *long-term memory* and *the writing process*. The *task environment* contains everything that is not the writer—including the writing assignment (*i.e.* topic and intended audience), and also external factors that effect the writer's motivation. The *task environment* also includes any text produced so far.

The model (shown in figure 2.5) decomposes the writing process into three broad sub-processes: *planning*, *translating* and *reviewing*. *Planning* develops the writing plan, *translating* converts the plan into text, and *reviewing* involves reading the text and editing.

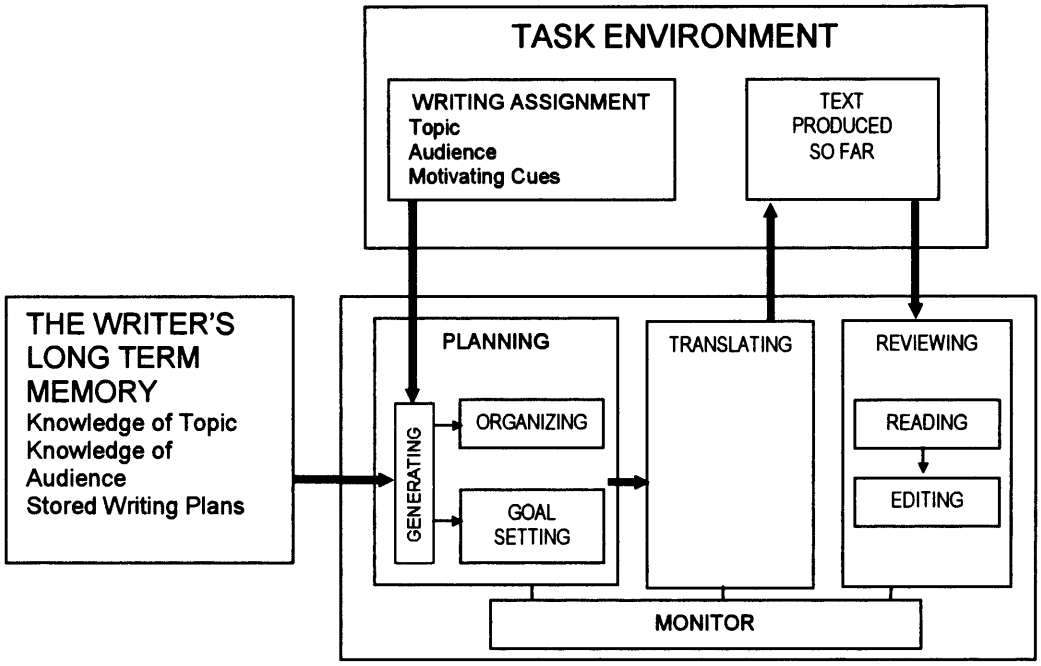


Figure 2.5 Flower and Hayes (1980) model of writing

Planning is decomposed into *generating*, *organising* and *goal-setting*. *Generating* is triggered by information about the topic and target audience (from the task environment) to retrieve information from long-term memory, including topic information or writing knowledge. *Organizing* then arranges the most useful items into a plan. Plans may be chronological, hierarchical, or both. Variations in *goal-setting* rules determine the writing styles:

- *Depth-first* (produce a perfect first sentence, then a perfect second sentence ... *etc.*);
- *Get it down as you think, then review;*
- *Perfect first draft;*
- *Breadth-first* (a draft is planned and then written out in full before review takes place);

Translating converts plans into text, and *reviewing* involves reading the text, and editing. Editing corrects violations in writing conventions (including stylistic issues), inaccuracies in meaning and the implementation of writing plans. Where significant editing is required the whole writing process can be recursively invoked. Editing may also occur in brief episodes, interrupting other ongoing processes. Finally, the *monitor* controls process calls through a set of production rules.

Although, from the perspective of the model, topic information arises entirely from long-term memory, the *generating* sub-process which manages this operation could conceivably be extended to incorporate external information-seeking. In the terms of the model, this would perhaps require a monitor production rule for invoking external information-seeking conditional on the requirement for content-based information which is not available in long-term memory. However, whilst such an extension might adequately describe the situation for precise, well-defined needs, care would need to be taken to ensure that it would satisfactorily accommodate information need uncertainty.

2.5.2 Writing as creative design - Sharples

In his 1996 model of writing as creative design, Sharples develops the idea of the writer as a designer of text. Central to Sharples' account (and particularly consistent with CSE) is the idea that a writer generates new material, and manages the abundance of possible next actions by imposing appropriate constraints. Sharples argues that the starting point of any writing process is the application of constraints that are either external, such as an essay topic, or internal, such as schemas, inter-related concepts, genres, and knowledge of language. As writing continues, the writer develops and monitors the text in terms of constraints and, if they deviate, brings them into harmony either by revising the text *or* by revising the plans. Success in writing (related to expertise)

comes from the way in which the writer is able to invoke just the right schemas and text structures to realise his or her goals.

Sharples explores the idea of linguistic knowledge as a constraint. He describes grammar as a generative system providing a framework for language production onto which a writer imposes schemas of knowledge and rhetorical structures appropriate to the task and the audience. He cites Scardamalia and Bereiter (1987) who describe writing as the interaction between two problem spaces—content and rhetoric. The content space is determined by the writer's beliefs about the topic, and the rhetorical space is determined by the writer's knowledge of text and writing goals.

At the highest level, Sharples model (shown in figure 2.6.) describes writing as alternating between two main processes: *reflection* and *engagement. Reflection* is decomposed into the sequence: *reviewing - contemplation - planning.* During *engagement* the writer is entirely devoted to the task of turning ideas into text. *Reviewing* involves reading what has been written and performing minor edits. *Contemplation* consists of knowledge exploration, calling up trains of associations and idea forming. And finally *planning* takes the results of contemplation and forms plans for their implementation.

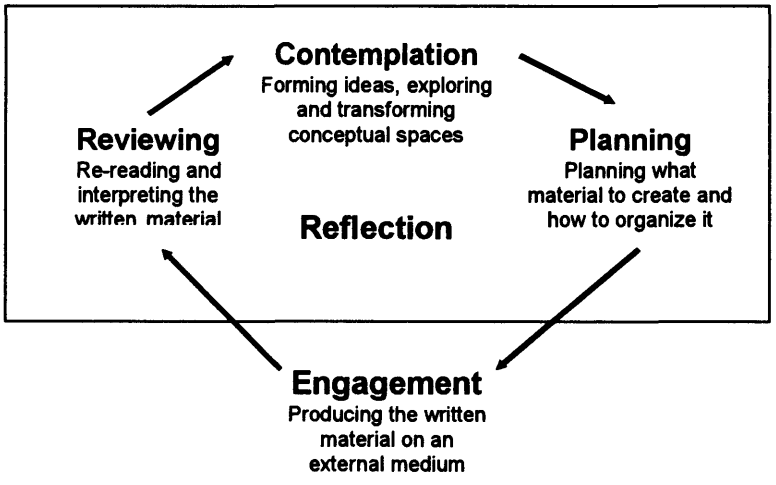


Figure 2.6 Sharples' (1996) model of writing

There are many similarities between Sharples' model and the Flower and Hayes model. The three major processes in Flower and Hayes' model: *planning*, *translating*, and *reviewing*, also appear in Sharples' model (*translating* becomes *engagement* in Sharples), with the minor modification that Sharples combines

planning and *reviewing* into a single macro-process of *reflection.* The main difference is that Sharples introduces *contemplation* within *reflection* for which there is no equivalent in Flower and Hayes. This addition reflects Sharples' emphasis on writing as creativity by including activities associated with creative thinking, such as the exploration and transformation of conceptual spaces. There is also a notable difference in model semantics. In Sharples' model arrows between processes represent sequence and information transfer between processes is implicit. In Flower and Hayes' model, arrows represent information flow and sequence is represented by the *monitor*'s production rules. Consequently, the Sharples model has no equivalent for the *monitor.*

The model by Hayes and Flower and the model by Sharples are similar insofar as, in broad terms, each describes *plan, execute* and *review* processes; they differ mainly in emphasis with Sharples' being more concerned with framing writing as a form of creative design. This perspective is useful and to some extent anticipates chapter 5 in which research into the psychology of design is used to develop a theoretical perspective on writing tasks which will be used as a backdrop for the interpretation of information-seeking phenomena.

Similar to Flower and Hayes, Sharples was primarily concerned with the idea of content being sourced from long-term memory, although, as with the Flower and Hayes model, the component which manages long-term memory access, which in Sharples' model is *contemplation*, provides a plausible process from which information-seeking might be managed.

2.5.3 Writing research in HCI

Research into the design and use of tools for writing in HCI has looked at a range of issues. To some extent efforts have been fragmented. Some work has considered how people interact with different kinds of input modalities for writing. For example, some have explored the design of pen-based input to text editors (for example, Thomas, 1987; Isokoski, 2001), or compared writing using computers to ordinary pen and paper (for example, Haas, 1989; Kellogg & Mueller, 1993). There has also been a strong thread of work in the CSCW community concerned with tools to support collaborative writing (for example, Neuwirth *et al*, 1994; Mitchell, Posner & Baecker, 1995; and Haake & Wilson, 1992).

Work that has focussed on the design of software considered from the perspective of the individual writer has tended to find its inspiration from a lack in traditional text editors for the creation and fluid manipulation of ideas as semi-discrete, modular macro-elements. Contributions to this approach have tended to differ from each other in terms of how the macro-elements and possible relationships have been conceptualised. For example, whilst Walker (1988) is relatively agnostic, referring to the representation of 'modular elements', others have aimed to support writing (and other tasks) through the representation of ideas as semantic networks (Halasz, Moran & Trigg, 1987), and writing specifically through the representation of documents as hierarchical structures (King & Leung, 1994), built from combinations of discourse elements (Schuler & Smith, 1990; O'Malley and Sharples, 1986).

O'Malley and Sharples (1986), for example, set out a framework and design proposal for a 'writers assistant'. Motivating their design was the differentiation of writing strategies such as brainstorming, following a thread (the generation of embedded subtopics), constructing an argument, drafting-redrafting, planning-drafting-revising. This variation suggested the need to enable a range of approaches to writing.

One approach that their system would support would be to allow the user to begin writing by creating discrete notes which could represent an idea at any level of complexity, from an indicative header to a fully formed piece of text. The user would be able to form a hierarchical network of notes (represented graphically) using discourse level relations such as *next, elaboration, example, definition, meta-comment* and *comparison* or simply leave notes unlinked. Alternatively, the user could begin by generating a stream of words. The system would then use surface features such as spacing and punctuation to automatically decompose the text into a constituent structure. And as the writer performs operations so the system would continuously monitor for constraint violations, such as spelling errors, wordy prose or word repetition.

Two points are raised here in relation to research into writing and writing tools. First, the user and the writing tasks they perform have been considered in rather generic, non-specific terms. There are perhaps historical reasons for this which relate to the relatively recent emergence of a more grounded, holistic and therefore contextually specific research mandate. The models discussed have

either been intuited or largely founded on theoretical insights rather than systematic observations of what writers do.

Second, from the perspective of writing, information seeking and information behaviour in general are embedded processes. Conversely, from the perspective of information seeking and information behaviour, writing is a common motivating activity. However, in both the models of the writing process and in research into the design and effects of different technologies on writing, the role of information-seeking and related information behaviour is hardly represented at all. The writer is predominantly seen as the primary originator of information. Perhaps this is because the availability of online databases, creating new opportunities for closely integrating information seeking within writing processes, is still a relatively new phenomenon. Whatever the case, the absence of both a contextualised, grounded perspective and the absence of information seeking behaviour as a component part of the writing process creates the opportunity for the grounded study of the role of information seeking and information behaviour in writing.

2.6 Summary and discussion

The story of information-seeking and, more generally, information behaviour research can be seen as a series of counterpoints around which a wide spectrum of approaches have evolved. Initially, system-centred IR research in the form of the Cranfield Paradigm provided a referent from which user-centred information-seeking research was to differentiate itself. The emergence of the Cognitive Viewpoint in user-centred information-seeking research provides a counterpoint for more holistic perspectives which make claims of greater contextual sensitivity.

Within user-centred research, Paisley criticised the lack of richness in information needs studies including the failure to consider the motivation for information-seeking and the uses to which it will be put. Paisley called for more holistic approaches and this was echoed by Dervin and Nilan who observed the early emergence of a new paradigm which was attempting to address these concerns. Dervin and Nilan also recommended a shift towards exploring the cognitive processes at work during information-seeking and the Cognitive Viewpoint emerged with this specific aim in mind.

The Cognitive Viewpoint concerns itself with understanding IR from the perspective of the cognitive structures of the people involved and has been championed in particular by De Mey, Ingwersen, and Belkin and Oddy. Taylor's levels of need largely falls within this viewpoint, as does Belkin, Oddy and Brook's ASK hypothesis; both focus on the uncertainty that information seekers experience in articulating their needs.

Some perspectives have emerged which have been less explicitly centred around cognition and more focussed on situational, behavioural and holistic issues. These include Dervin's Sense-making, Ellis' and Bates' Behavioural Approach, Kuhlthau's ISP theory, Wilson's Information Behaviour theory and the work of Byström and Järvelin, Vakkari and others. In these approaches, cognitive structures are de-emphasised in favour of issues such as communication, information behaviour, affect and the role of users' wider tasks in determining and structuring information behaviour.

Dervin's sense-making is concerned with the user's perspective and the situations that they find themselves in for which information is needed to make sense of some aspect of their world. Sense-making studies focus on eliciting aspects of the SITUATIONS-GAPS-USES model often using the Micro-Moment Time-Line Interview. The Behavioural Approach centres on information behaviour usually framed within a rich view of the information user's task situation. Within this approach, Ellis has provided a contextually located and yet generalisable taxonomy of prototypical information-seeking behaviour characteristics based. This includes: *chaining*, *browsing*, *differentiating*, *monitoring* and *extracting*. Also behavioural in character, Bates' Berrypicking model emphasises information need evolution through the systematic exposure to information during the information-seeking process.

Kuhlthau's ISP model locates information-seeking within task goals and is concerned with behaviour and affect in addition to cognition. The ISP model incorporates notions of uncertainty and evolution through the movement from vague ideas to a focussed perspective. Wilson, who has been a particularly strong advocate of holistic research in Information Science, broadened the unit of analysis to information behaviour in general and integrated a number of models within his own which treats wider tasks as problem solving and divides

them into stages, each corresponding to the resolution and reduction of uncertainty. Uncertainty, and the idea of the user's wider task being framed as problem solving, appear in Byström and Järvelin framework of task complexity as corresponding with *a priori* determinability, and Kuhlthau's idea of task formulation being a pivotal point in information-seeking was recognised and explored further by Byström and Järvelin, Vakkari and others. Finally, and specifically in the domain of journalism, Nicholas and Martin have provided an analysis of journalists' information needs according to Lines' original framework under which information 'function/purpose' serves to contextualise needs. They identified five purposes; these were: *fact-checking*, *current-awareness*, *researching*, *to obtain context* and *stimulus.*

The tension between the holistic and cognitive approaches centres around the non-cognitivist view that the Cognitive Viewpoint fails to deal adequately with issues of information behaviour context. For the non-cognitivist, the debate revolves around a distinction between inner and outer worlds as a source of interpretation and explanation. It is argued that context, whether physical or social or both, plays an essential role in shaping human behaviour and that the Cognitive Viewpoint fails to account for this. In defending the Cognitive Viewpoint, though, Ingwersen argued that it imposes no theoretical limitations for "exploding the system of environmental variables" (Ingwersen, 1992, p.19). For Ingwersen, the information seeker's social and organisational context are indeed within the scope of the Cognitive Viewpoint insofar as they transform internal cognitive structures.

In essence, Ingwersen's response is correct. Whilst the Cognitive Viewpoint concerns itself predominantly with what goes on in the head, what goes on in the head is the representation and processing of the user's situation. This includes their situation within an organisation, a department, a social, ethnic or religious group etc., and of course within a physical environment. Hence, contextual factors can, and indeed must fall within the cognitive approach. Further, aspects of context which are important for explaining interaction are *precisely* those that are internalised and understood by the user. The inner/outer distinction, when invoked as a criticism of the Cognitive Viewpoint, is misleading. The information seeker's inner world is essentially *about* their outer world.

The Cognitive Viewpoint, then, does necessarily not focus on inner life at the expense of outer life, but this is not to say it has, as it happens, dealt with context adequately; it is just to say that *in principle* it could. The style of the cognitive approach is to develop accounts that impose structural formality on cognitive representations and processes. Formality has power since it requires precise definition, clear examples and well understood structures. But, given the complexity of real-world situations and the complex understanding that people need in order to negotiate these, representations of a person's situation in terms of cognitive structures can at best be impoverished (Anderson *et al.,* 1993) and at worst intractable. If we are to embrace context in all its complexity, then the Cognitive Viewpoint is unlikely to offer a manageable way of doing so.

The emergence of Situated Action, Distributed Cognition and Cognitive Systems Engineering in HCI are also developments away from a strictly in-the-head cognitive approach. Suchman argued that plans and indeed goals are intrinsically vague and do not in a strong sense determine behaviour; rather, we are continuously walking into social and physical situations to which we must respond. This view provides a particularly appropriate way of understanding information behaviour. Suchman, however, also claimed that the meaning of behaviour can only be found through an understanding of the context in which it occurs and so cannot be descriptively separated from it. But it was argued that some selective separation must occur and that this provides the foundation for abstraction and generalisation.

Distributed Cognition, whilst taking a cognitive perspective, is also contextual insofar as it seeks to soften the inner/outer divide by regarding cognition as extending outside of the head. According to this view, problem solving can be distributed across members of a social group, can involve the coordination between internal and external (material or environmental) structure, and can be distributed through time such that products of earlier events can transform the nature of later events.

CSE is a holistic framework for modelling work systems to inform the design of technology. Where Suchman's Situated Action makes no commitment to which contextual factors might be important for interpreting and explaining human behaviour, in the context of complex work tasks CSE focuses on modelling behaviour as situated within a set of active constraints and a set of available

resources. These, it is argued, whether they be cognitive, cultural or material circumstances, shape choices and structure behaviour and so should be the focus for analysis.

Both the Information Science and the HCI literature have evolved approaches which argue for the importance of understanding the context of behaviour for interpreting and explaining it. Writing news reports and feature articles in a newspaper is the task context for the information behaviour studied in this thesis and the review considered two models of writing taken from the Cognitive Science literature and reflected on how they might incorporate information-seeking from external sources. Hayes and Flower's model, which is relatively formal, decomposes writing into: *planning,* which develops a writing plan, *translating*, which converts the plan into text, and *reviewing*, which involves reading the text and editing. A monitor controls process calls using production rules. Sharples' model, which emphasises the role of constraints in narrowing the writer's problem space, is broadly similar to Hayes and Flower's except that Sharples adds a *contemplation* process which reflects his emphasis on representing writing as a creative process.

Finally, research into the writing process and HCI research into tools to support the writing process have tended to be generic in nature and have omitted to consider the role of information seeking within that process. With online databases now available to professional writers as intrinsic and essential part of the information landscape, there is now a need to consider how writing and information behaviour mutually condition and shape each other.

The aims of this chapter were to review literature for two reasons: the first, was to provide a perspective within which the general meta-theoretical approach of the thesis could be located. The second was to review ideas and findings that would act as reference points later in the thesis. In terms of meta-theoretical approach, the integration between information behaviour and its context, the consequent priority placed by many researchers on studying information behaviour in context, and the holistic approaches that this entails have been considered. Research question 1 asked:

What are newspaper journalists' prototypical information behaviours in relation to the seeking and use of information from electronic news

cuttings services whilst writing news reports and feature articles, *and* what are the aspects of their task situation that explains them?

This question most closely associates the thesis with the Behavioural Approach and in particular work of Ellis. Like Ellis, the question takes a holistic perspective in order to contextualise accounts of information behaviour within its task context. In chapters 3 and 4 in particular, studies are presented which attempt to determine information behaviour during news report writing and explain these in terms of the wider task context. In chapter 4, the CSE view of work activity as bounded by constraints and resources demarcates those aspects of context of interest and this provides structure to a Grounded Theory analysis of interviews with journalists. The importance of understanding the constraints under which people work was stated by Paisley (1967), is intrinsic to CSE and was explicitly adopted by Sharples (1996) in his model of writing; in chapter 5, a constraint based interpretation of complex tasks provides the theoretical basis for a design-psychology oriented explanation of information behaviour within writing tasks.

The question of providing reference points for issues raised later in the thesis related most closely to research question 2, which asks:

How does this knowledge [developed from question 1] **relate and contribute to more generalisable theory of information behaviour in relation to the processes and structure of complex information tasks?**

This chapter has considered a number of models and findings which will arise elsewhere in the thesis. In particular, chapters 3 and 4 will relate findings to Ellis' Behavioural model (1989 & 1989b), Bates' Berrypicking model (1989), findings by Nicholas and Martin (1997), and Suchman's Situated Action (1987). In chapter 5, ideas from design psychology are related to concepts of uncertainty in Belkin, Oddy & Brooks' ASK hypothesis (1982a, 1982b), Kuhlthau's ISP model (1993), notions of focus formulation from Kuhlthau's ISP model and later related work by Vakkari *et al.* concerning focus formulation (Vakkari, 2000a; Vakkari 2000b; Vakkari & Hakala, 2000; Vakkari & Pennanen, 2001; summarised in Vakkari, 2001). Chapter 5 also deals with notions of task complexity developed by Byström and Järvelin (1995) and the studies by Yang (1997) and Tang and Solomon (1998).

Chapter 3

Orientation: A lab-based exploratory study

3.1 Introduction

The study reported in this chapter begins to address the first of the research questions stated in chapter 1 by exploring the kinds of information behaviour that occur during a news writing assignment. Since little is known about how news writing and information behaviour relate it was decided to begin the empirical work with a qualitative, exploratory study in order to provide some initial perspectives which might form areas of focus to be developed through subsequent studies. In effect, this approach is equivalent to the Grounded Theory technique of theoretical sampling (Strauss & Corbin, 1990). Hence, this study acts as an exploratory precursor to the field study carried out with journalists at *The Times* reported in chapter 4.

The exploratory study performs a detailed, lab-based examination of a small number of subjects performing an invented news writing assignment using electronic information resources including news cuttings services. The emphasis is not on drawing a few, firm conclusions from statistical data, but rather on exposing some issues for further consideration within a rich view of someone, as Paisley recommended, "at work, under constraints and pressures, creating products, drawing upon the elaborate communication network that connects him with sources of necessary knowledge" (Paisley, 1968, p.3). Breaking news stories often arrive on a news reporter's desk in the form of an agency newswire, and it is the journalist's job to reinterpret this information and exploit the many information sources at their disposal in order to construct a news report. Electronic resources, particularly ENC archives, play a key role in this process.

The exploratory study had two broad aims: the first was to describe global patterns of information behaviour and writing activity as these occur in relation to each other. To provide structure to this aim the study uses a framework synthesised from a framework for electronic information-seeking by Marchionini (1995) and a framework embedded in Sharples model of writing (Sharples, 1996). The second aim was to bring to the fore local behaviours (at a lower level of description) which may be worth examination in further studies. The criterion for this selection was relatively loose, being only that from the perspective of usability, current technological solutions appear poorly suited, and benefits

might be realised from a more informed design. To illustrate this with an example (and anticipate the study findings slightly), subjects showed that whilst writing they would occasionally want, or anticipate wanting, to re-consult a document which they had read on-screen but which would later, after many other documents had subsequently been viewed, be difficult to recover.

The structure of the remainder of this chapter is as follows. The next section describes the study set-up in detail, including details of the subjects, their tasks, the materials that were provided and how data was recorded. In section 3.3, the analysis method is also described. Two basic approaches to data analysis were used (coding and memoing) in relation to global and local foci. Section 3.4 details the findings and section 3.5 provides a discussion.

3.2 Method

The subjects were non-journalist PhD subjects recruited from the Department of Computer Science at UCL. In the study, which was performed in a laboratory setting, screen recordings and talk-aloud protocol recordings were taken of subjects writing a mock newspaper news report based on information given to them in a fictitious newswire release and information that they would find by searching web resources including a selection of news cuttings services. The study was intended to replicate a typical newspaper news assignment. Of course, a reporter given such a task is likely to seek information through many information channels, but given that the focus of the study was the use of electronic information services, and in particular ENC services, these were the only information resources made available.

The study set out to examine the performance of 2 subjects[1] in detail, although, given some early subjects' low use of online resources (a prerequisite for information behaviour) and a lack of detail in their verbal protocols, five subjects were tested in all with the final two being used in the analysis. This procedure was considered acceptable since the study was not intended to verify an existing hypothesis or model but as an exploratory exercise.

[1] This may appear to be few subjects, but it is actually 1 more than in Flower and Hayes (1980) seminal paper on the cognition of writing.

Data were recorded in the form of digital screen capture movies and synchronised audio recordings. Before the task, subjects were asked to 'talk-aloud' and were prompted during the task where appropriate. To encourage verbalisation, and also to ensure that any differences in search ability would not affect subjects' ability to perform the task, the task was arranged as a negotiated, collaborative effort with the researcher acting as search intermediary. Hence, subjects were not given direct access to the information resources, but were forced to communicate requests and strategies to the researcher on all aspects of searching and browsing. The researcher's computer was running a web browser with its homepage set to a simple, custom-built portal linking to potentially useful resources including ENC services and various other search resources such as generic web search engines. A screen shot of the portal, which includes descriptions of the linked resources, is shown in figure 3.1.

Resource	Description	Date range	Help pages
Financial Times Global Archive	A searchable archive of world newspapers, trade publications, magazines and newsletters, wire services and others. Sources are selectable. Note: searches will cover the last 3 months unless you refine the date range.	5 years	Search tips on the search page
Guardian Unlimited	Search the Guardian and Observer archive by keywords in headline or headline and text and coarse date range.	Sept 1998 to current	Search tips on the search page
ITN news archive	Search the archive of ITN news stories by keywords, section and coarse date range.	About 6 months	none
BBC News archive	Search the archive of BBC news stories by keywords in headline or headline and text, section and coarse date range.	Nov 1997 to current	Help Tips
Hansard	Search the archive of transcripts of House of Commons debates by keywords, speaker and fine date range.	Nov 93 to current	Help page
Google	Search the worldwide web by keyword.	-	Help page
Altavista	Search the worldwide web by keyword, date and language.	-	Help page
Excite-UK	Search UK or European web sites by keyword.	-	General search tips Advanced search tips

Figure 3.1 A screen shot of the custom portal used as a starting place for the subjects' task.

Subjects were provided with direct access to a computer running a word processor for making notes and for writing their report. Each of the two computers was connected to a second slave monitor positioned so that both researcher and subject could see each others' screen. The physical arrangement of the study is shown in figure 3.2.

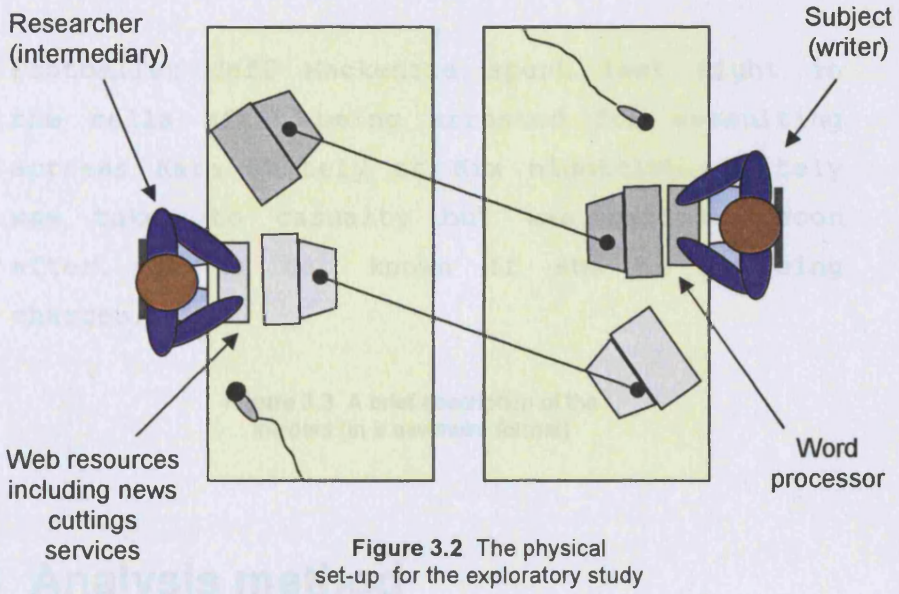


Figure 3.2 The physical set-up for the exploratory study

Given that separate computers were used for searching and writing, it was not possible for subjects to gather information as they might have done using clipboard based copying, nor were subjects able to request printouts. However, it *was* possible for them to record information manually by typing notes into their word processor file. They were also able to request that entire documents be retained by opening them in a new browser window to be kept open while searching continued in the original window.

Subjects were given a fairly detailed task scenario in which they were asked to imagine that they were a reporter working for a national tabloid newspaper and had been assigned to report a nightclub incident occurring between two moderately well-known celebrities, a footballer and an actress. The celebrities were real but the incident was not, being invented for the purposes of the study. The incident involved the footballer assaulting the actress and was designed to be at least plausible.[2]

Subjects were handed their assignments in the form of written instructions (shown in full in appendix I). This contained a brief description of the incident (in a newswire format) and instructed them to write a 300 to 400-word report

[2] Some dates, names and roles have been changed in order to protect the patently innocent.

elaborating the story using background information. The newswire report is reproduced in figure 3.3.

> Footballer Jeff Mackenzie spent last night in the cells after being arrested for assaulting actress Kate Whitely at Kix nightclub. Whitely was taken to casualty but was released soon after. It is not known if she is pressing charges.

Figure 3.3 A brief description of the incident (in a newswire format)

3.3 Analysis method

From the movies transcriptions were produced which included the subject and intermediary verbalisations, descriptions of the subject and intermediary interactions with their respective systems (including descriptions of the subjects' evolving notes and reports) and system responses.

3.3.1 Analysing for global patterns of activity

Human task performance can be considered at many level of description ranging from a macroscopic view of a task as a whole down to a detailed account of individual actions. Analysis of this study was performed at two complementary levels; the first was an analysis of global patterns of activity throughout the task as a whole, and the second was a more focussed examination of individual behaviours. The purpose of the global patterns analysis was to identify large-scale behaviour patterns and also to provide a context within which the more detailed analysis of individual phenomena could be understood.

Once created, the transcripts were segmented according to transitions between subject verbalisations, intermediary verbalisations, subject interactions, intermediary interactions, system state changes, and also according to pauses in activity of more than a second or two. The segments were then coded according a framework which was developed broadly as synthesis between Marchionini's (1995) model of the sub-processes involved in information retrieval

and Sharples' (1996) process model of writing. These were used since they provide a convenient decomposition of behavioural sub-processes that form part of interaction with information retrieval systems and writing.

The codes used, along with their counterparts from Marchionini and Sharples are shown in table 3.1. Where it was considered appropriate, some of the Marchionini and Sharples process elements were decomposed into lower-level constituent processes.

Exploratory study codes	***Codes from Sharples (1996)***
Writing report	Engagement
Editing report	
Reading report	Reviewing
Developing report structure plan	Planning
Developing notes	
	Codes from Marchionini (1995)
(no equivalent)	Extracts Info
Finds 'useful' information	*(no equivalent)*
Reading/browsing/searching document	Examine Results
Select Document	
Browsing results list	
Submit Query	Execute Query
Composing/editing query	Formulate Query
Select information resource	Select Source
Information-seeking goal articulation	Define Problem
(no equivalent)	Recognize Accept

Table 3.1 The codes used for the global patterns analysis, with their counterparts from Marchionini (1995) and Sharples (1996)

In addition, a cursory look at the data suggested that expressions of information need were not always immediately followed by their attempted resolution; rather, information goals were often deferred. In order to capture the temporal relationships between need expression and attempted resolution, the subjects' expressions of information need (sometimes verbal, sometimes written in note form) were also coded, each with an individuating label.

The labelling of expressions of information need was complicated by the fact that one information-seeking goal might subsume another. For example, an

information need expression could be fairly general, such as 'find information about x', whilst another on the same topic could be more specific, such as 'find out the age of x'. Further, a general need expression can conceal or evolve into a number of more specific forms (for example, through further consideration/articulation by the user), and so it was important to devise a coding scheme that could relate the different forms together since this would be necessary in order to formally capture all need deferrals.

A rule was devised that would relate information need expressions on the same topic articulated at different levels of specificity. The rule, called *the subsumption rule*, was that if the set of all information relevant to need expression *a* is also relevant to need expression *b,* but some of the information relevant to need expression *b* is not relevant to need expression *a,* then need expression *a* can be said to be subsumed by (*i.e.* is a specialisation of) need expression *b*. The situation that would satisfy the subsumption rule is represented in figure 3.4 using Boolean set notation. a^I and b^I represent the sets of information relevant to a and b respectively.

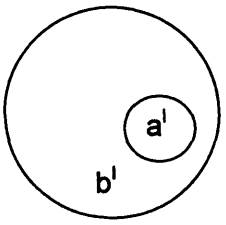


Figure 3.4 A Boolean set representation of the situation in which the subsumption rule is satisfied

Using the rule, subsumption relations were coded using a hierarchical coding scheme. First, each expression of information need was labelled with a letter denoting a general theme. Needs of the most general form ('find information about S') where S is the theme, were labelled with a designated theme letter (uppercase). Where an expression of information need was subsumed by a more general need it would be additionally qualified using a lowercase suffix *e.g.* 'Sa'. For each theme, suffixes were allocated alphabetically according to order of occurrence within the protocol. The suffixing strategy was also applied recursively, so, for example, if expression 'Find out about P's playing history' was classified as 'Sa', the more specific expression 'Find out when P moved from team A to team B' would be sub-classified 'Saa'. Figure 3.5 gives an

example of how an extract of the resulting hierarchy might be structured around a theme 'S'.

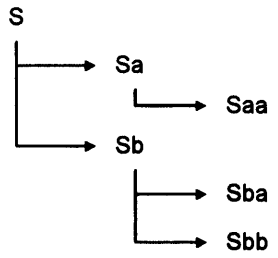


Figure 3.5 An example extract of a hierarchy representing need subsumption relations

3.3.2 Analysing for local behaviours

In addition to the identification of global patterns of activity, the analysis sought to explore individual phenomena which might potentially have implications for the design of integrated electronic information-seeking and writing tools. Since it was not known *a priori* what kinds of issues these might be, it was decided that this would be best served by an informal analytic approach. The approach used was to attach memos to protocol segments and sequences of segments as appropriate, to categorise similar events, and then to report these as narratives based on selected examples taken from the transcripts.

3.4 Findings

Complete reports produced by subjects 1 and 2 are shown in appendices IIa and IIb respectively.

3.4.3 Global patterns of activity

Following coding, a visual representation was created for each subject's data showing the coded activities as they occurred through time. These take the form of grids which will be referred to as *activity timeline grids*. The activity timeline grids are shown for subjects 1 and 2 in figures 3.6 and 3.7 respectively. In the grids, each column represents a minute-long time period and each row represents an activity type according to the coding scheme. Where a given activity occurred at least once within a given minute the corresponding cell is

filled black. A black cell does not indicate that the entire minute comprised of that activity, but rather that it was recorded at least once during that minute. Consequently, for any given minute many cells may be filled.

Sections are also added to the bottom of each grid to indicate when information needs were articulated and when they were subsequently pursued. Articulation points are shown as labelled boxes with their horizontal position corresponding to the time of articulation (Vertical offset is used only as a means of visual separation and has no semantic significance). The box labelling scheme follows the need subsumption hierarchy. Each box then connects with an arrow pointing into a horizontal bar above, which shows when information retrieval activities were performed which were judged as corresponding to that information need being pursued (selecting information resources, composing/editing queries, submitting queries, browsing results lists *etc.*). Vertical black lines within the bar show general changes of theme and grey vertical lines represent encounters with information considered useful by the subject (white areas in the bars represent periods of information retrieval without useful information being found).

The ordering of activities in the grids from top to bottom (shown in the left-most column) is intended to broadly reflect the idea of activities serving as preparation for or facilitating another (the one above). For example, before you can *compose or edit a query* (second from bottom), you must have *selected an information resource* (bottom), likewise before you can *submit a query* (third from bottom) you must have *composed or edited a query* (second from bottom). *Writing*, being the ultimate aim of the exercise, appears at the top of the list. Also, according to this scheme, *developing notes* and *developing report structure plan* appear below those activities more directly related to creating a piece of text (*writing*, *reading* and *editing*) and information-seeking activities appear below *note-taking* and *planning*. Hence, broadly speaking, where any given activity cannot be performed, the researcher/writer must cascade down until they find a lower level preparatory activity that *can* be performed. This will then by followed by systematic climbing, with activities at each level facilitating those at the level above. If this preparation idea is correct, then cascading and climbing should be visible in the way that activities are distributed across the activity grid.

Four observations were made on the basis of the activity grids. These were:

1. An initial phase of intense searching and note-taking occurred prior to a clear-cut switch to writing interspersed by gradual increases in report reading and editing.
2. Despite initial intense searching though, searching can be reinitiated during latter stages of the task.
3. The ordering of activities does appear to broadly correspond with the preparation idea as evidenced by cascading and climbing.
4. Information needs *were* occasionally deferred, and this appeared to be so that attention and effort could be maintained on a current activity. This could be writing a complete draft, or pursuing information needs relating to a different theme.

3.4.3.1 An initial phase of intense searching and note-taking and then a clear-cut switch to writing, reading and editing.

The time-line grids for both subjects show an initial period of intense information-seeking and note-taking. This was followed by a clear-cut switch to writing and gradual increases in report reading and editing. Subject 1 began with a 28 minute period of searching and note-taking followed by a central 26-minute period of intense writing with some reading and editing. For subject 2, the switch from information-seeking and note-taking to writing, reading and editing occurred around minute 70.

Notably, the changes from the initial phase into writing, reading and editing were clear-cut, and the talk-aloud protocols show that these corresponded with explicitly articulated decisions.

> **Subject 1:**
>
> (min 24) I reckon that's the searching over. [reviews the information he had gathered]
>
> (min 28) That's my notes. I reckon I'm gonna move to the document.
>
> **Subject 2:**
>
> (min 67) OK. So, I'm gonna start writing this thing now.

During the initial phase, the subjects compiled their notes partly from extracts of information that they had retrieved, but, equally often, they both also noted down information recalled from memory (*i.e.* brainstorming). In this sense, their notes appeared to act not just as a way of recording material, but also as a way of

organising important information into a single information resource, perhaps with more stable recall characteristics than both long-term memory and online-archives (*i.e.* recall is guaranteed).

3.4.3.2 Searching can be reinitiated in the latter stages of the task

Despite the initial intense searching period, both subjects returned to searching after writing had been started. For subject 1 this happened after he had written a complete first draft (hence writing had ended) and during a final phase of reading and editing. For subject 2, searching was reinitiated as writing dissipated more gradually. This difference appears stylistic; according to the styles identified by Hayes and Flower (1980) subject 1's strategy was to *get it down as you think, then review* whilst subject 2's strategy appeared closer to the *perfect-first-draft* approach.

Where information-seeking is motivated by a wider task, it is often considered as a preparatory activity. The observation that searching was reinitiated despite an initial period of information-seeking does not necessarily demonstrate this not being the case for the subjects in this study; information-seeking, whenever it was performed, was performed in support of the writing task, and so in this sense was always preparatory (Indeed, if the wider task could have been adequately performed without any information-seeking, it undoubtedly would). What the re-initiation of searching does show, however, is that even though the subjects may have attempted to anticipate their information needs in advance of engaging with writing their reports (suggested by the initial intense period) new needs nevertheless arose throughout.

3.4.3.3 Cascading-down/climbing-up

The ordering of activities appears to broadly correspond with the preparation ordering idea of cascading and climbing. For both subjects, activity began (left) by dominating the lower regions of the grid, and ended (right) by dominating the upper regions. Also, throughout the grids a series of minor diagonal sequences indicating local periods of climbing are visible. A close succession of diagonal and vertical sequences indicates periodic climbing interspersed by rapid downward cascading and the re-initiation of climbing. In fact the vertical blocks that appear in the grids invariably correspond with the same pattern, although, given the resolution of the grids, climbing is too rapid to discern.

Subject 1

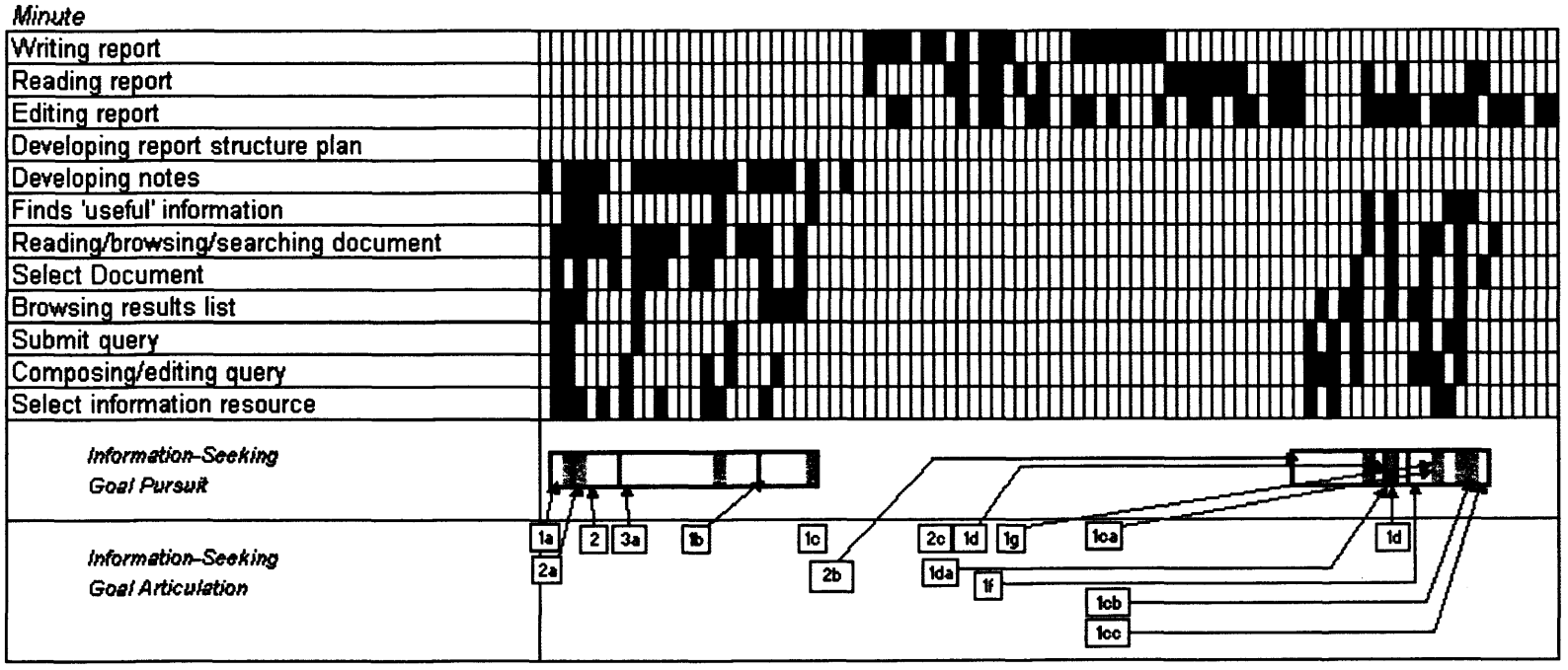


Figure 3.6 The activity timeline grid for subject 1 showing macro-patterns of activity according to the coding framework.

Subject 2

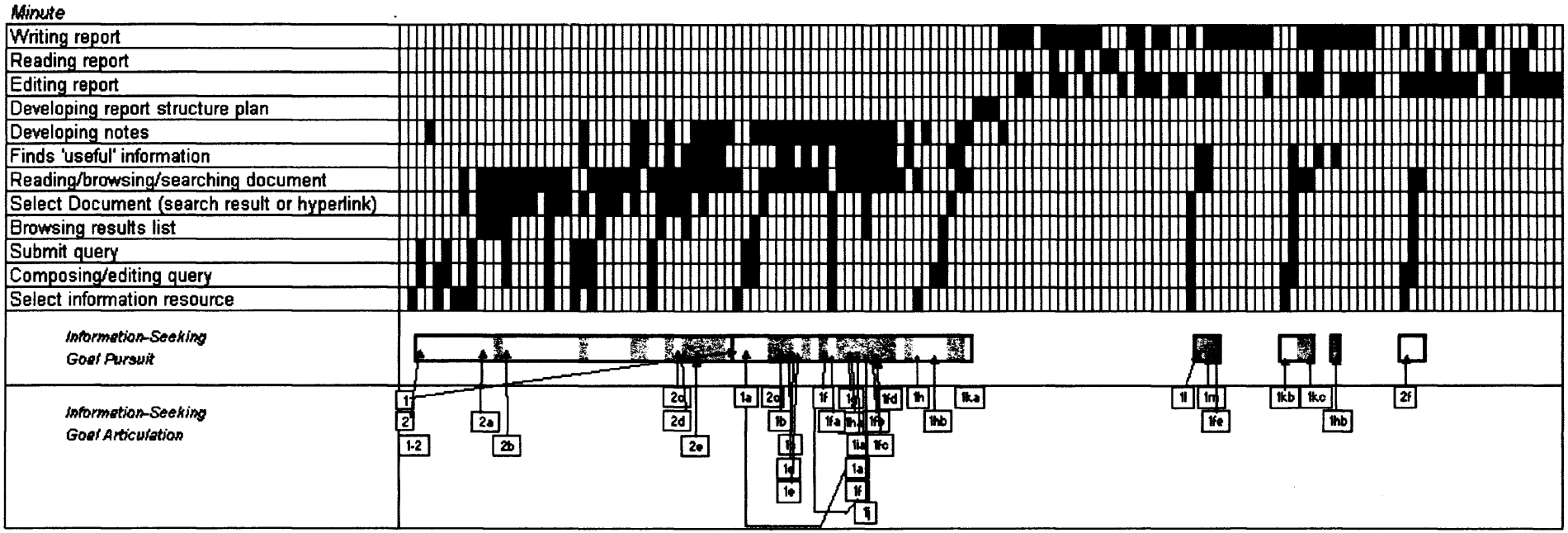


Figure 3.7 The activity timeline grid for subject 2 showing macro-patterns of activity according to the coding framework.

3.4.3.4 Information need deferral

The activity timeline grids for both subjects show many occasions when information-needs were articulated but activities for resolving them were deferred until later. (Need articulation points are indicated in the grids by boxed numbers. Arrows leading away from these boxes indicate when they were addressed.) Deferrals of this kind are particularly evident for subject 1; for example, he expressed need 2b in minute 25 during a review of the information he had gathered:

> **Subject 1:** (min 25) She was 91 in the.. Most beautiful women in the wor... I think she was.. She was in the LM's most beautiful women in the world... '96' or something. I think she was {unintelligible} ... Erm... So we might have to go and search a bit more detail about that.

But he did not address this need until minute 67. In total he deferred 10 out of 16 information needs.

In some cases, subject 1 deferred information needs so that he could address another need which he had articulated at about the same time. For example, in minute 1 he articulated needs '1a' and '2a' and chose to turn attention first to need '1a':

> **Subject 1:** (min 1) Alright. Erm... Yeah. Can you get me erm... the ages of Jeff Mackenzie and Kate Whitely? ...erm ...probably be able to find it ...erm Let's have a look. I suppose you just search Google for "Jeff Mackenzie" and come up with a football site.

He then returned to need '2a' in minute 4:

> **Subject 1:** (min 4) OK, and yeah, can you do the same for Kate Whitely? Probably on Google as well.

For the most part, though, subject 1's deferrals occurred in a central period of intense writing, reading and editing. During this time he deferred all occurring information needs, noting them down in his draft report as he wrote to ensure that he didn't forget them. An extract from his report at minute 49 shows some examples:

> The football career that shot him into the public eye during the mid-nineties now lies in tatters. He was one of football's hottest properties when City signed him from United for a (check price - and was it a British record?) back in ????. Although his spell at City brought moderate success (?)
>
> (**Subject 1:** report extract, min 49)

Compare this to the equivalent section in his final report:

> The football career that shot Mackenzie into the public eye during the mid-nineties now lies in tatters. He was one of football's hottest properties when City signed him from United for a British transfer record back in *1996*, but had already acquired a 'bad boy' tag, following rumours of dressing room unrest. Despite a moderately successful spell with City, ...
>
> (**Subject 1**: report extract, final version)

In total, more needs occurred to subject 1 while he was writing (8) than during the initial search period (7). Also, a comparison of the distribution of letter suffixes used to identify the information needs shows that the needs that occurred while he was writing were generally more specific (indicated by the number of suffixes) than those that occurred earlier.

An explanation for subject 1's deferrals during the central writing period relates to his *get-it-down-as-you-think-then-review* strategy. As he was writing, and unanticipated information needs occurred to him, he was faced with a choice; he could either stop writing and address each need as it occurred, or he could note it

down and maintain the flow of writing. In keeping with the strategy he chose the latter.

However, when the time came he did not choose to address all the needs he had written down. In minute 35 he wrote:

> Kate Whitely, like Justine, is one of television's hottest properties. She stats [starred] in the hit BBC drama Rescue, alongside stars such as ??????? and ???????, who will be shocked to hear .....
>
> (**Subject 1**: report extract, min 35)

But during minute 59, when reviewing his report and addressing needs he had noted down, he changed his mind:

> **Subject 1:** (min 59) No one cares about who stars in Rescue alongside Kate Whitely. They're not particularly big names. And, and of course they're gonna be shocked... to hear.... so its like... whatever...

Subject 2, only deferred one information need (out of a total of 35), being happier to interrupt the flow of writing in order to return to information-seeking. The one deferral occurred early in the task when he expressed an interest in finding information about both Jeff Mackenzie (the footballer) and Kate Whitely (the actress). These expressions are represented in figure 3.7 by the need articulation boxes '1' and '2' respectively. (The box marked '1-2' represents the goal of finding out how they knew each other. In reality they didn't know each other, and when this was pointed out he abandoned this need). Since he couldn't address both needs at the same time, subject 2 first pursued need 2 (and subsumed variants '2a', '2b', '2c', '2d' and '2e') and then, after 37 minutes, returned to need '1'.

In summary, both subjects occasionally chose to defer information-seeking in favour of alternative competing goals—competing, that is, for limited resources. For both subjects, the competing goals included other, often simultaneously occurring

information needs. For subject 1, though, they were additionally, and most often, writing goals which were chosen over information-seeking in order to maintain the flow of writing.

3.4.4 Local behaviours

The identification of lower-level behaviours was based on whether, from a usability perspective, current technological solutions appear poorly suited. The behaviours are organised around 4 headings:

1. Biography seeking
2. Quotation seeking
3. Confirming proper name spellings
4. Information-gathering

3.4.4.1 Biography seeking

Subject 2 performed extended periods of information-seeking based around the goal (which he articulated explicitly) of finding biographical information about both key people involved in the incident. First:

> **Subject2:** (min 13) Erm... erm... I want to find out... some sort of biography on her. Lets try biography, so lets try 'Kate Whitely AND biography'

And later:

> **Subject2:** (min 40) So let's go to... Jeff is bound to be in the papers and stuff isn't he. Erm... Who's going to have the best biography on it?... We want a reasonable length one really. We want to look back quite far. BBC news archive's not bad is it.

At that time subject 2 related his desire for a biography in terms of a need for general information:

> **Subject2:** (min 41) Erm.. We want sort of general information. That's why its kind of biographical stuff I'm looking for.

The search resources portal used in the study offered links to news archive sites and Web search engines, but none of these provided means for specifically requesting biographical documents. Consequently, it was not possible for subject 2 to achieve a good mapping between his need (as he conceptualised it) and any query that he could produce; in Taylor's (1968) terminology, his need had to be compromised. Subject 2 performed many searches over a period of 30 minutes looking for biographies but he didn't find any. However, during this time he was able to gather a good deal of biographical information from the various sources that he encountered.

3.4.4.2 Quotation seeking

Another information-seeking goal that occurred during the tasks which was poorly supported by the query mechanisms used was finding quotations. Both subjects spent time searching for quotations—in one case a quotation was known to exist, in another its existence was speculated.

Subject 1 indicated the need for a known quotation in the notes within his first draft:

> Mackenzie, who has recently shown himself to be recovering from the illness that 'drove him to the edge' (CHECK THIS ON BBC SITE)...
>
> (**Subject 1:** report extract, min 39)

Having some knowledge of football and the footballer in question, subject 1 had recalled something that the footballer had been quoted as saying which, through writing the report, he realised he would like to include. Later, while reviewing his report, he tried to find it and explained:

> **Subject 1:** (min 77) I wanna... I wanna quote, what he said about.... he never said that he'd commit suicide, but he said, its something like it drove him to the edge and he might not be here today...

The strategy that he chose to find the quote was to search for all stories about the footballer on an ENC service and then to browse the results with a rough knowledge of where the quotation would appear in terms of the chronology of events and dates. He was unable to find the quotation. His response was then to alter his original text to something less "specific" (subject 1, min 78) , *i.e.*:

> Mackenzie, who has recently shown signs of recovering from the depression that he claims almost caused him to take his own life...
>
> (**Subject 1:** report extract, min 78)

Subject 2 also wanted to find a quotation by the footballer. He expressed this while he was reviewing a results list as part of his initial information-seeking period (*i.e.* before writing) after he read a headline that the footballer had spent some time helping children in a school following an incident of bad behaviour. Identifying this as potentially relevant background information, he asked to view the article and said:

> **Subject 2:** (min 47) It would be better to have a quote for him on this though wouldn't it. It'd be nice to get that

No quotations were found in the report, but by now the idea of finding a quotation had become more important to him and he asked to see another report about the same story. He also had an idea of the gist of the quotation that he was looking for:

> **Subject 2:** (min 48) Er, we'll get a quote out of that... A quote of him saying, "I've been naughty, but I'm not really naughty" or something. That'd be good.

But he could only speculate that it might exist. He then found a candidate quotation:

> **Subject 2:** (min 48) Ah, brilliant. [reading] "Maybe in this circumstance, I'll hold my hands up and say 'Naivety', but at the end of the day I haven't killed anybody" [laughs] Oh yes!

He typed this into his notes and later included it in his report.

For both subjects, then, finding past quotations by particular people with a particular gist was important but because the search tools didn't provide means for specifically locating quotations their default strategy was to browse reports related to particular events.

3.4.4.3 Confirming proper name spellings

During his intensive writing period, subject 1 added into his report the prediction that, given the incident being reported, the footballer would now very likely face the sack from his team manager. He wrote the manager's name, but then added a note to himself to check the spelling:

> Mackenzie, who has recently shown himself to be recovering from the depression that 'drove him to the edge' (CHECK THIS ON BBC SITE) following his acrimonious departure from Rovers to Leicester. is almost certain to be given his marching orders by new boss, Wanderers' David Roane (check spelling).
>
> (**Subject 1:** report extract, min 40)

Thirty-five minutes later, when he was reviewing his text, he addressed this need using the strategy of constructing a query for a generic web search engine using an assumed spelling. He then reviewed a few results from this search and confirmed that they related to the person.

Confirming name spellings also emerged as a concern for subject 2, although on both occasions this happened while he was reading and making notes from source documents that contained the required names and so he was able to check them quickly.

3.4.4.4 Information-gathering

In the study, the means provided for information-gathering were limited; given that searching was performed on a separate computer from writing, it was not possible to use clipboard supported copying. It was also not possible to print documents. Given the amount of information-gathering, some frustration was expressed about the lack of convenient functionality:

> **Subject 2:** (min 28) Er... I can't cut and paste can I? Which is quite annoying.

To gather information it was necessary for the subjects to retype it manually, or, to open a document in a separate browser instance which would be kept open. Subject 1 requested this, but in fact, didn't look at the document again:

> **Subject 1:** (min 21) Yep, ooh yeah actually. Just keep this page up. This is the one that I want, er... For most of the story actually.

Despite the clear advantage of gathering information at the time it is encountered, at various times both subjects wanted to return to documents that had been displayed previously in order to gather specific information. For both subjects this was prompted by writing about what had been in these documents in their reports.

In minute 90 subject 2 was writing about the incident that had led to the footballer spending time helping in a school, when he remarked that he wanted to include the date that it had happened but had not noted it down. He then asked the intermediary to relocate the original document.

Similarly, in minute 104 subject 2 was writing about an incident between the footballer and a previous girlfriend when he realised that he had omitted to note down her name:

> **Subject 2** (min 104) OK, so, where's this other girlfriend? Don't we have her name? I don't think I do. Can't you find out that older girlfriend's name? Er.. Maybe, where did we find it originally? It was either football unlimited or BBC wouldn't it have been? Let's try BBC.

Requests to return to source documents appeared to happen for one of three reasons:-

a. *Extraction omission* - The subject read some information he considered valuable but failed to note it down;

b. *Extraction error* - The subject read some information he considered valuable but miss-recorded it;

c. *New information goals* – At some point in the task the subject evolved a new information need for which some previously read document was considered useful;

3.5 Discussion

The objective of this study was to provide an initial perspective on information behaviour in relation to electronic resources in the context of a writing task, in order to establish some areas of focus to be taken up and developed further through work described in the rest of the thesis. Given the focus on journalistic research and writing, a news reporting scenario was used. The study highlighted a number of issues which were divided into *global patterns* and *local behaviours*.

3.5.5 Global patterns

The identification of global patterns of activity was made on the basis of the *activity timeline grids* which reflected the application of a coding framework to the data. The activity grids showed an initial period of intense information-seeking and note-taking followed by a clear-cut switch to writing and gradual increases in report reading and editing. However, both subjects returned to searching after writing had been started (and in one case finished). Also, the structure of activity was described as a pattern of 'cascading and climbing' through various levels of preparation both at a local and a global level.

A link between these findings and potential design considerations is perhaps not immediately obvious. If phases of searching/gathering and writing/reading/editing were more categorically defined then, for example, it might have been useful to consider the resources used during each stage and to investigate varying their prominence at the user interface to reflect task stages. However, the possibility of information needs occurring throughout the task, and in particular being prompted by writing itself, mitigate against this as a valuable direction.

The third pattern was the deferral of information-seeking to a time other than when the originating need occurred. Deferral was most evident in the activities of subject 1, and in his case many information needs were deferred in order to maintain the flow of writing. Despite information needs arising during writing, there was a reluctance to allow these to distract attention away from it. Information needs also occurred during writing for subject 2, although he chose to interrupt writing to resolve them.

The wish to maintain attention and avoid interruptions is perhaps a familiar experience when performing complex tasks. Sullivan (1993) explains this phenomenon in terms of what he refers to as 'cognitive momentum'[3] (see also

[3] There appears to be no connection between 'Cognitive Momentum' as used by Sullivan, and concepts of the same name used by Donaldson (1999) to mean a kind of involuntary pattern

McFarlane, 1998). Sullivan argues that the disruption caused by interruptions during complex task performance arises because in order to perform such tasks, people must 'align' a number of cognitive resources so that they will coordinate and cooperate with each other. Further, establishing the necessary alignment requires effort, and the alignment can be disrupted by interruption; hence interruptions carry a cognitive overhead. The situation is analogous to a manufacturing scenario in which resources, in the form of people and machinery, are individually and collectively configured to perform a given kind of work. Further, given that the setting-up costs are normally independent of the amount of work that is then done by the production line, mass production proves an economical option.

The desire to maintain 'cognitive momentum' when writing in the face of unanticipated information needs, and the strategy of creating 'fact gaps', suggest that integrated information retrieval and writing systems ought to allow users to record emergent information needs quickly and easily so that momentum can be maintained. Of course, as evidenced by subject 1, a text editor (in conjunction with an improvised representational convention) provides an ideal resource for this with the benefit that it can preserve the proposed location and any intended resolution strategies. Hence, on its own, the observation of need deferral provides little additional leverage for functionality requirements for integrated information retrieval and writing systems beyond an already essential integrated text editor.

3.5.5.1 Local behaviours

The local behaviours identified in the study were *biography seeking*, *quotation seeking*, *confirming proper name spellings*, and *information-gathering*.

In discussing these, it seems important to address the extent to which they might be regarded as arising due to factors specific to the scenario used in the study, or due to properties of the task considered at some more general level of description. To what extent are the findings representative of research and writing tasks in general?

---

completion and by Miura (1996) to mean the adaptation and optimization of visual processing resources. In this thesis Sullivan's meaning is used.

To what extent are they representative of news research and writing tasks in general? And to what extent are they only representative of celebrity news research and writing tasks? In the discussion of local behaviours the generalisability issue is considered.

3.5.5.2 Biographical information needs

In the study, subject 2 explicitly prioritised finding biographical information but the search resources were poorly suited to this type of need. If he had found a biography on either of the two main characters in the incident, this probably would have significantly reduced his information-seeking time. Since it is possible to imagine a search engine operating over database which would make such a need relatively straightforward, this finding provides an issue for the design of integrated information retrieval and authoring systems at least in the domain under study; hence it will be explored further in the thesis. However, in the context of writing about two moderately well-known celebrities the need for biographical information is perhaps unsurprising. What if it had been to report a train accident or a business merger?

A biography provides a narrative account of someone's life detailing the most interesting and important events (ODLIS). Hence, a biography is a kind of overview. Further, the need for overview documents might generalise well into other kinds of news writing assignment (such as reporting train accidents or business mergers) and, for that matter, research and writing tasks in general. Hence, this finding will be taken forward into the field study by considering it as a need for overview documents. The questions that will be asked will be: Do journalists commonly have the need for overview documents on the topic they are writing about? How are these needs resolved with current systems? How might they be optimally resolved?

3.5.5.3 Quotation seeking

The examples of quotation-seeking in the study indicate the value that the subjects placed on using quotations as part of their reporting. In one case a subject was aware of the existence of a quotation, but failed to find it; in the other a subject hypothesised correctly that an appropriate quotation might exist and succeeded in

finding it. An interesting similarity is that in both cases the subjects were able to specify the required speaker and the essence of what the quotation should say, *i.e.* 'drove him (or me) to the edge' and 'I've been naughty, but I'm not *really* naughty'. The examples suggest that an ideal quotation search tool would allow users to search a collection of news reports using *speaker* and *gist* attributes as optional query parameters.

Searching for quotations in news story archives may be a goal that usefully extends beyond the current reporting scenario to news reporting and feature writing in general, and given that it is a goal which is poorly supported by generic keyword search engines, it may be worthy of further exploration in the context of the planned field study. The questions that are raised are: Does the goal of quotation-seeking within ENC services commonly arise during news report and feature article research and writing? If so, what constraints or attributes of a required quotation can be specified *e.g. speaker, subject matter, gist*? How might the sorts of quotation seeking that journalists perform be optimally supported?

3.5.5.4 Confirming proper name spellings

The results of this study suggest that finding out how to spell proper names correctly is a goal that arises as a result of the news reporting task. This may appear a relatively insignificant goal, but for a journalist, particularly one working on international assignments (and therefore dealing with unfamiliar names), to short deadlines and strict accuracy requirements, it may well be a very significant issue. Consequently the problem of spelling proper names will be pursued in the field study.

Naturally, standard word-processor spelling checkers don't help here. The strategy of searching a generic Web search engine using a number of spelling variations provided a solution for subject 1, and demonstrated the relatively resourceful adaptation of an available tool for solving a problem that was undoubtedly not anticipated by the tool's designers. Adaptations of a similar type have been observed elsewhere, in particular by researchers studying aviation and ship

navigation (Hollan, Hutchins & Kirsh, 2000) and represent an interesting interaction issue. For one thing, it is important for designers to be aware of such incidental, adaptive uses since this can not only permit improvements which better accommodate the use, but not being aware of them can result in 'improvements' that design such use out of a tool altogether.

The questions for the field study concerning proper name spelling will be: Do journalists researching and writing news reports and feature articles identify confirming proper name spellings as a frequent goal? If so, how do they achieve this goal? How might this goal be optimally supported?

3.5.5.5 Information-gathering

Information-gathering has been observed and discussed by many researchers as a part of information-seeking (see for example, Ellis (1989a), and Marchionini (1995)), and received particular emphasis as a part of Bates' Berrypicking model (1986). By gathering information encountered during searches, users essentially construct a resource to support their writing. They anticipate needing to find information again easily and so they select information that they encounter and put it in an accessible location; in doing so they generate an easily locatable distillation of what they regard as useful - a document with a higher density of relevant information than any other single document that they will work with in the assignment. Gathering information can be thought in terms of Distributed Cognition in the sense that the products of some actions (finding information) can transform the nature of later events (writing) in a way that offloads demands on working memory into the environment *i.e.* the writer doesn't need to remember everything that they find.

Despite its importance as a part of information-seeking, information-gathering has received little attention as an issue for information systems design. It is clearly an issue for systems that attempt to integrate information retrieval services with writing tools. The results of this study suggest that the need to transfer information from source documents into a task specific collection is a frequent activity in the task of interest and so this will provide a further focus for the subsequent field study.

Specifically, the questions that we will ask will be: How do journalists transfer information into a task specific repository (and consequently what form does this repository take)? How might information transfers be better supported?

Although they were provided with tools to record information as it was encountered, when they came to write, both subjects wanted to re-find source documents. Sometimes users can anticipate needing information and so they gather it, but, sometimes they omit to gather information that would be judged useful at the time (extraction omission), miss-record information (extraction error), or simply change their minds about some information long after they have finished viewing a document (new information goals), and these can lead to the difficult and time-consuming task of reconstructing past searches. The ease with which a user can relocate previously seen documents, then, is clearly a matter for system design. Browser history lists allow users to backtrack through seen documents, but the visual representation of these lists is often cryptic, difficult to access and difficult to relate to the task activities that led to the information being viewed. Browser 'back' buttons, use a linear navigation paradigm which only stores the current navigation branch (*i.e.* cul-de-sacs are unobtainable).

The need to relocate previously seen documents is a phenomenon which is not accounted for in standard berrypicking theory and, if it is a goal characteristic of online research and writing tasks, then it is one that integrated information retrieval and authoring systems should accommodate with minimum user-effort; consequently it will be pursued in the subsequent field study. The questions will be: Do journalists find that they need to refer back to previously seen source documents? If so, why does this happen–to what extent is it due to *extraction omissions, new information goals* and *extraction errors*? How might relocating source documents be better facilitated?

3.5.6 The situated writer

In places through this discussion findings from the current study have been related to some other models and theoretical perspectives *viz.* cognitive momentum, berrypicking and Distributed Cognition. The discussion will conclude by relating a number of the findings to ideas from Suchman's Situated Action (Suchman, 1987).

In what was referred to in chapter 1 as the *vague plan argument,* Suchman (1987) challenged what she regarded as a piece of social science orthodoxy. She referred to this orthodoxy as the 'planning model' and her challenge was that the model overemphasised the role of plans as determinants of human action; that few actions are actually explicitly planned at all and where they are, plans are inherently vague. For Suchman, the efficiency of plans comes from the fact that they do not represent action in all their concrete detail, but rather that they provide a high level orientation, after which responding to the contingencies of real life situations as they unfold is managed by the *ad hoc* and improvised application of "embodied skills" (Suchman, 1987, p51). A closely related point that she makes is that often it is only through engaging with a situation that its possibilities become clear, and so we are in fact not in a position to know the outcome of our activities in any detail when they begin.

In terms of the current study, these ideas provide a valuable explanatory context through which a number of the findings can be understood. Both subjects preceded writing with a period of preparation. This period, which we can regard as orientating, involved seeking information and gathering the information they anticipated they would need when writing. Hence, during this period the subjects made judgements about what they would write in their final report. In this sense, gathered information is a representation of intent. But despite this preparatory phase, the subjects were unable to anticipate new information needs arising once they had engaged with writing—a time when information-seeking was less convenient (evidenced by the deferrals). And some of these needs resulted in the desire to go back and gather information from documents that had already been viewed.

On the one hand, the occurrence of the preparatory information-seeking phase suggests a pre-planned component to the writing. On the other, information needs occurring during writing suggest a more reactive control mode in which the subjects responded dynamically to their own evolving texts. As they committed words to screen, so they found themselves with unanticipated information needs. One very good example of this is when subject 1 wrote:

> Mackenzie, who has recently shown himself to be recovering from the illness that 'drove him to the edge' ... (CHECK THIS ON BBC SITE)...
>
> (**Subject 1:** report extract, min 39)

The example suggests that introducing Mackenzie's illness into the text prompted the recall of a quotation, on recalling it the decision was made to include it, and this required that the wording be checked. It is as if, through engaging with writing, the text evolved a direction which was both guided and responded to by the subject.

Findings like this suggest that research and writing tasks of the type investigated in this study involve an interplay between both planned and situated action. In Suchman's terms, the tasks were situations that the subjects "walked into"' (Suchman, 1987, p54) with uncertainty not only about what they would find out but a related uncertainty about what they would ultimately produce. The brief given provided only outline to their goals which then evolved as a function of their interaction with the situation. Early decisions concerning useful content were preparatory estimates, and only through the act of writing itself did they come to decide precisely what they would say.

This chapter began with the objective of beginning to address research question 1. Research question 1 is concerned with identifying and explaining prototypical information behaviours in relation to electronic news cuttings services that occur when people write news reports and feature articles. The role of the study was to perform a qualitative, exploratory examination of the kinds of information behaviour

that might occur in order to provide some initial areas of focus to be developed through the field study which is reported in the next chapter. The exploratory study was a detailed, lab-based examination of a small number of subjects performing an invented news writing task using electronic information resources including news cuttings services. It had two broad aims: first, to describe global patterns of information behaviour and writing activity as these occur in relation to each other, and second to identify more local behaviours, occurring at a lower level of description.

Globally, the study identified 4 patterns. First, initial intense searching and gathering occurred before a clear-cut switch to writing intermingled with gradual increases in report reading and editing. Second, despite this switch, searching could be reinitiated during the latter stages of the task. Third, the ordering of activities broadly corresponded with a cascading and climbing pattern. Fourth, information needs were occasionally deferred so that attention and effort could be maintained on a current activity. However, a link between these patterns and novel design requirements was not immediately obvious and so these will not be pursued in the field study.

In terms of local behaviours, the study identified *biography seeking* (which was generalised to *seeking an overview*), *quotation seeking*, *confirming proper name spellings* and *information-gathering.* The latter included relocating documents following an initial viewing to gather further information. Each of these issues was related to requirements for novel, specialised functionality and so will be considered further through the field study. Also, some phenomena identified in the study, collectively demonstrated that the subjects' performance of the task was characteristic of Situated Action as conceived by Suchman.

Chapter 4

Uncertainty and change: A contextually rich model of newspaper journalist's information behaviour

4.1 Introduction

This chapter reports an *in vivo* interview study (also reported in Attfield & Dowell (2003)) of the information behaviour of a group of journalists at *The Times* newspaper in London. The study develops on research question 1 by further identifying journalists' prototypical information behaviours in relation to ENC services, and confirming those identified in the exploratory study. The study seeks to explain information behaviours in terms of the journalists' task situation. The chapter also addresses research question 2 by developing a model of the journalists' information behaviour and by relating the study findings to extant theory.

In the previous chapter, some issues were developed based on findings from the exploratory study which were to be addressed through the field study. These related to the need to find overview documents, the need to find quotations, the need to confirm proper name spellings, the need to gather information, and also to refer back to previously seen documents. The study also highlighted the Situated Action idea of vague plans and goal specifications as providing a means for interpreting the observed behaviour. These issues provide some focus for the current study, although given the opportunities presented by an *in vivo* study for uncovering additional issues, it was decided that the field study, whilst carrying forward issues identified by the exploratory study, would not be overly constrained by them. Consequently, whilst investigating the issues raised in the exploratory study, the field study was additionally intended to take a broad perspective with the aim of providing data for a more general model of journalists' information behaviour.

The field study used a qualitative methodology based around the techniques and procedures of Grounded Theory, broadly in accordance with Strauss and Corbin (1998). As a generative, emergent research methodology, Grounded Theory is particularly well-suited for developing holistic, contextually rich models and is appropriate for studies in HCI and Information Science where the aim is to make explicit the processes at work within complex situations. For the study, Grounded Theory was adapted so that it might incorporate some key concepts taken from Rasmussen, Pejtersen and Goodstein's (1994) Cognitive Systems

Engineering (CSE) approach. In particular, the idea of understanding activity as structured and determined by constraints and resources plays an important role in the analysis and provides structure to the resulting model. The use of a framework to structure a Grounded Theory analysis, however, might be regarded as compromising the Grounded Theory tenet that concepts should emerge unconstrained from the data. This question will be dealt with in the chapter.

The structure of the chapter is as follows: the next section (4.2), discusses Grounded Theory, CSE, and the way in which these were combined. Section 4.3 describes the study method in detail, including descriptions of participants, interviews and specifics of the analysis method. Section 4.4 reports the findings in outline and in detail. The findings in detail are presented in sections according to the constraints-resources-behaviour framework as numbered sub-sections. Finally, in section 4.5 the findings are summarised and then discussed in relation to Bates' (1989a) Berrypicking model, Ellis' (1989) Behavioural model, Nicholas and Martin's (1997) assessment of journalists' information needs, and Suchman's (1987) notion of Situated Action.

4.2 Approach: Combining Grounded Theory and Cognitive Systems Engineering

Grounded Theory is a set of analysis techniques originally developed by Glaser and Strauss (1967) which provides a means for creating theory from qualitative data. Central to Grounded Theory is the idea of coding—the linking of phenomena with conceptual labels. Using coding, data is initially fragmented from its original form (*e.g.* interview transcripts) and then reconstituted in terms of underlying concepts and relations (*i.e.* theory). Grounded theory uses three kinds of coding: *open coding*, *axial coding* and *selective coding*. Although the analyst will typically alternate between these techniques depending on the needs of the analysis, producing a Grounded Theory generally begins with open coding and ends with selective coding.

In open coding, labels are attached to instances of phenomena in the data. The phenomena might be events, actions, states or objects. Abstraction is at the heart of the method, and is achieved by grouping similar phenomena into

higher-order 'categories'. The associations made through open coding classify phenomena and categories into higher-order categories, and this forms one of two parts of a Grounded Theory. For the second part, *synthetic* (typically causal) associations are identified between categories through axial coding. Where a number of categories are linked to a single category through axial links, the many are designated *sub-categories* (this should be differentiated from the type subordination established through open coding).

The third coding technique, selective coding, draws out a single 'core' category around which all the major categories are organised, and by integrating subsidiary categories around the core category is used to develop a descriptive 'story line'. In a sense, selective coding is a process of prioritisation in which some aspects of the data may be disregarded if they do not integrate well into a single focussed perspective.

CSE (Rasmussen, Pejtersen, & Goodstein, 1994) was described in chapter 2 as a framework for modelling work to inform technological system design. CSE aims to model complex socio-technical work systems in order to predict how people would behave in response to engineered changes—to ask, 'what could be done differently and better'? Central to CSE is the idea that complex, dynamic work does not have predetermined procedures. Understanding actor discretion is particularly important for understanding the trajectory of complex information work, and this makes this kind of work activity unnameable to traditional task analysis which is more suited to local, stable task procedures with few, well-defined resources. CSE recommends that the analyst systematically build a picture of a person's 'action alternatives' (Rasmussen, Pejtersen, & Goodstein, 1994) by modelling the boundaries of available choices. According to the view, choices are bounded and hence shaped by active *constraints* and available *resources* and these are what the analyst needs to refer to when explaining activity. Hence, "the crucial question is not "what" the actors are doing, but "why", together with the alternatives for "how"." (Rasmussen, Pejtersen, & Goodstein, 1994, p31).

The 'constraints and resources' perspective is a framework which focuses attention on particular kinds of issue. Using a framework has the effect of prescribing what kinds of phenomena will be sought. Hence it determines what kinds of phenomena will be 'seen' (and potentially missed), and so by using a

framework, an entirely 'open' approach is compromised. However, in promoting their approach Rasmussen, Pejtersen and Goodstein (1994) argue that a well-defined point-of-view is must be established in order to achieve rapid convergence. The question of whether or not to use a framework, then, involves a trade-off between openness to the data on the one hand, and efficiency on the other. Clearly the latter choice, though, must rest on the judgement that the framework of choice is well suited to the goals of the study.

Using the idea of constraints and resources as explanatory concepts within an otherwise Grounded Theory approach impacts on the types of phenomena and categories identified (through open-coding) and on the types of relationship found (through axial coding). In fact, in Grounded Theory itself, there is some precedence for the use of *a priori* frameworks. Whilst the original conception of Grounded Theory was that relationships (and consequently the phenomena related) should emerge unconstrained from the data, Strauss and his latter colleague Corbin proposed a framework to guide axial coding which they referred to as *the paradigm model* (Strauss and Corbin, 1998). Disagreement about the legitimacy of this is at the heart of the distinction between the versions of Grounded Theory separately promoted by Glaser and Strauss (Dey, 1999), with Glaser suggesting multiple framework alternatives, each as valid as each other. Nevertheless, in Strauss and Corbin's approach is the recognition of the value of providing a framework as a tool for helping the analyst think "systematically" (Strauss and Corbin ,1990, p.99).

The question of whether to use a framework depends upon the clarity of the aims of the study and the appropriateness of the framework. If the kinds of phenomena and relationships of interest can be declared *a priori*, an *a priori* framework is suitable. Further, it is argued that for the purposes of identifying and explaining information behaviour in a work context, the questions to be asked are essentially the 'what', 'why' and 'how' suggested by CSE. Consequently, in the context of the openness/efficiency trade-off, the concepts of *activities*, *constraints* and *resources* were used in this study in order to focus open and axial coding.

4.3 Method

4.3.1 Interviews and participants

Data in this study were gathered primarily through semi-structured interviews with twenty-five journalists at *The Times* in London. The participants were 19 Home News Reporters, 4 Feature Writers, 1 Obituary Writer and 1 Chief Sub-editor. Levels of experience ranged from 1 to 36 years. Some additional data were gathered through email correspondence. Interviews were conducted at the participants workplaces and lasted between 20 and 40 minutes.

Interviews typically began with the researcher prompting the journalist to give a general description of the process of working on a news or feature assignment. During the interview the researcher would steer the conversation towards issues of information-seeking and information use from ENC archives during the process. Specific questions relating to the behaviours identified for investigation during the exploratory study (seeking overview documents, quotations and proper name spellings, information-gathering and referring back to documents) were asked where the opportunity arose. The interviews were recorded.

After 14 interviews, the recordings were transcribed and an initial analysis conducted. The initial analysis allowed more specific questions to be formed with the intention that they would become the focus for a further period of interviewing (*i.e.* selective sampling). These questions were then used to produce a questionnaire around which subsequent interviews could be structured.

The questionnaire (shown in appendix IIIa) provided a well-structured questioning procedure developed with the aim of increasing density within the existing model around axial coding (described in 4.3.3) using the coding framework that had been developed in relation to understanding the 'what', 'why' and 'how' suggested by CSE. During the early stages of this further set of interviews, however, it became apparent that the use of a highly structured approach had a negative effect on the flow of data gathering predominantly due to a disconcerting effect it had on interviewees. Consequently, the interviews returned to a more conversational style in which the particular issues addressed by the questionnaire were pursued where possible. Hence, the coverage of

issues in an interview depended on how well the researcher was able to exploit and direct the flow of conversation. Following the second round of interviews, the new recordings were transcribed and the analysis developed further. An sample interview transcript is shown in appendix IIIb.

4.3.2 Open Coding

Open coding involves developing concepts and categories in order to provide the structure of the theory. Given the 'constraints and resources' framework, open coding was used to identify three basic types of category: *activities*, *constraints* and *resources*. Properties and dimensions were used to define and differentiate categories. The approach taken was to focus the identification of dimensions and properties on comparative category differentiation.

Within each broad category type (*activities*, *constraints* and *resources*) a strict approach to property inheritance was taken. Each category was associated with dimensions and defined in terms of values on those dimensions with the possibility of any dimension remaining unspecified. Subordinate categories inherited the same set of dimensions as their superordinates with the addition of at least one new dimension value. This could mean that a previously unbound dimension acquired a definite value, or an existing value was further qualified. Hence, from bottom to top, categories became more abstract.

4.3.3 Axial Coding

Axial coding is the making of links between categories in terms of their *synthetic* (typically causal) relations. The constraints and resources framework formed the basis for axial coding in this study. Hence, the study pursued the *whys* and the *hows* of each activity in these terms. Due to an apparent bias towards modelling adverse events in a clinical setting, Strauss and Corbin's (1990) paradigm model was not used.

Given the logic of abstraction central to Grounded Theory subordinate inheritance should apply not just to definitional dimensions and their values but also to axial relations (the *whys* and the *hows*). For example, where the rationale for travelling to work is to change location from home to work, so the rationale for travelling to work by bicycle is also to change location from home to work but also (for example) to get some exercise. Similarly, where the *means* for learning

how to configure a printer is to read the printer manual, so the means for installing a printer's cartridges is to read the printer manual section about installing cartridges. In each case, subordinate categories inherit axial relations and additionally specify them. In practice, this strict approach to inheritance proved valuable for achieving integration and density in the analysis and for sense-checking the type hierarchy and axial links. If an axial relation could not be meaningfully inherited by a subordinate category, it was a cue for reviewing the superordinate/subordinate status.

4.3.4 Selective coding

The role of selective coding is to identify one category which appears to represent the central phenomenon of the study from those developed during analysis. This category is given priority, perhaps at the expense of one or two competing categories. It was considered that an exercise in requirements capture ought to strive to be more comprehensive than this strategy would allow. What is important in software design is not the prioritisation of one concern above all others, however novel or insightful, but the balancing of multiple competing design considerations into an integrated solution. Evolving a single theme at the expense of all others might well result in a reductionist solution in which due consideration was not given to multiple factors. Hence, priority was not placed on developing a central theme at the expense of other potentially valuable parts of the data.

4.4 Findings

4.4.1 The high level view

Figure 4.1 shows the high level *analytic* organisation (or taxonomy) of the categories resulting from the analysis. The full taxonomy is shown in appendix IIIc. The taxonomy content and structure reflects the priority given to the constraints and resources framework. Below the high-level representation, section 3 (activities) is deeply structured, subsuming many types and sub-types. Axial links connect activities with relevant constraints and resources. Categories subsumed under categories 1 and 2 of the model effectively provide the context for the activities described under category 3.

The findings are described in detail in the next sub-section, organised in terms of the category taxonomy. Process is a very important aspect of the phenomena being examined here as it is for Grounded Theory analysis in general, and this is not brought out by the taxonomy structure. However, process will be explored as the account progresses.

Category 1 Constraints
- Category 1.1 Angle
- Category 1.2 Deadline
- Category 1.3 Word-count
- Category 1.4 Prior written commitments
- Category 1.5 Constraints on content
  - Category 1.5.1 Newsworthiness constraints
  - Category 1.5.2 Historical context constraint
  - Category 1.5.3 Accuracy constraint
  - Category 1.5.4 Legal constraints
  - Category 1.5.5 Explanation constraint
- Category 1. 6 Constraints on structure
  - Category 1.6.1 Cut-from-bottom constraint
  - Category 1.6.2 Original wording constraint

Category 2 Information Resources
- Category 2.1 External information resources
  - Category 2.1.1 External read-only information resources
  - Category 2.1.2 External read-write information resources
- Category 2.2 Internal information resources
  - Category 2.2.1 Domain knowledge
  - Category 2.2.2 General writing knowledge
  - Category 2.2.3 Newspaper writing knowledge
  - Category 2.2.4 Resource knowledge
  - Category 2.2.3 Internal report plans

Category 3 Information Behaviours
- Category 3.1 Information-seeking
  - Category 3.1.1 Exclusivity checking
  - Category 3.1.2 Background information-seeking
  - Category 3.1.3 Seeking evidence for a hypothesis
  - Category 3.1.4 Information-seeking for feature comparison
  - Category 3.1.5 Confirming names and how to spell them
  - Category 3.1.6 Identifying useful contacts
- Category 3.2 Information-gathering
  - Category 3.2.1 Dragging and dropping
  - Category 3.2.2 Printing
- Category 3.3 Information reviewing
  - Category 3.3.1 Reviewing information gathered during an assignment
  - Category 3.3.2 Reviewing information read but not gathered during an assignment
  - Category 3.3.3 Reviewing information read prior to an assignment

Figure 4.1 The high level structure of the model.

4.4.2 Findings in detail

4.4.2.1 Category 1 Constraints

An important aspect of the findings from this study is that uncertainty and change permeate many aspects of the journalist's task. This provides an important perspective from which to interpret and understand the journalists' information behaviour. The sources of uncertainty and change can be traced to two issues: first, how well the task constraints are known at the outset of an assignment, and second, the extent to which they change during it. Since constraints provide structure to a task and determine what the tasks is; so as they reveal themselves or change, the task takes on new shape.

An approach was adopted in the analysis that would attempt to bring underspecification and change in the constraints to the fore. This was done by defining each constraint, seeking to single out elements within them that were observed as being subject to potential uncertainty or change, and by seeking to explain how this uncertainty or change might come about. These factors were identified as dimensions or properties for each constraint category.

The first dimension for each constraint is its *prescription*—a statement of what that constraint stipulates. The stipulation is expressed in general terms as it applies across assignments and does not differ from assignment to assignment. For example, the first constraint prescribes that each report should follow a specified angle. But this expression of the constraint is general inasmuch as it is independent of what the angle happens to be for any one assignment. What the constraint *means* to the journalist on any given occasion, and how it impacts on their subsequent activity, of course, depends on what the angle is. In a sense, the constraint expressed at the general level is incomplete or non-specific. The second dimension, under the title *determinant*, draws out from the *prescription* the factor which binds the constraint at any given time to a particular meaning, and so connects the constrain with specific activities; in this case, the angle itself. By knowing that assignments should follow an angle and by knowing, in a particular case, what that angle is, the journalist is able to proceed. For a given constraint there may be multiple determinants.

The significance of identifying the *determinant* within the *prescription* is not simply that it is the factor that changes from instance to instance (or assignment to assignment), but that this is the factor that is the source of uncertainty and change. In the example of the *angle* constraint, the angle can be (and often is) subject to mid-assignment change. To take another example, the *accuracy constraint* specifies that an article should report the truth accurately—this is the prescription. What binds this prescription for any one assignment is what the truth relating to a story is, hence the *determinant* is the truth. But, knowing what the truth is may not be a simple matter and may be a source of uncertainty (and consequent information-seeking). And as information is encountered through information-seeking or incoming newswire reports, so the facts as they are known can change, and this can alter the nature of the task to a greater or lesser extent. Hence, identifying the determinant is an exercise in locating a source of potential uncertainty and change.

Each constraint, then, has a prescription and a determinant dimension. The subcategories also list a *source.* This corresponds to the origin of the *determinant* from the writer's perspective. For example, in the case of the word count *determinant*, the *source* is editorial decision making. The *source* is the entity that establishes what the *determinant* will be in any one case and hence indicates where the journalist would need to look in order to discover it. The final subcategory is *cause of a priori indeterminability* which, where applicable, provides an explanation for why a *determinant* might be unknown or unknowable at any point in time.

The constraints which are included in the model are not exhaustive; to list all active constraints in a work situation would undoubtedly be a task without end. Rather, the constraints which are included are selected as those which will help in providing an explanatory context for the activities which appear later in the model.

Category 1.1 Angle

Each report should follow a specified angle—this is the *prescription* property of the *angle* constraint. Throughout the interviews, the idea of an *angle* featured highly in the participants accounts of their work, and the pursuit of a particular angle was described as being at the heart of every research and writing assignment. Every assignment angle had to be sanctioned by editorial staff

and would usually be an editorial decision (*source*) communicated to the journalist through a verbal briefing. However, the knowledge of more senior journalists enabled them to have a more or less significant input.

> Given the importance of the angle, it will be useful to explore in a little detail what the journalists meant by it, and how it affects their work. The angle was described as a "thrust" or "driving force" of an assignment; the new "twist" or "hook" or guiding idea that directs the story. The initial view of the researcher in the study was that an angle might be a kind of value judgment imposed by the journalist on their research and writing. However, on exploring the concept, it appeared that the angle took the more tangible form of a proposition, question or hypothesis.

The case for the angle as a hypothesis was made initially by some examples that were given. These included: "Were the lessons learnt?", "Were they implemented?", "If they were, why has this happened again?". When explaining that he would highlight and retain documents he received if he identified in them a possible angle that might be worth exploring later, a travel correspondent said:

> **BW:** "I am highlighting angles, possible angles, possible things that may or may... I mean often you know you start with a hypothesis that something might be the case and you go out and try and find out if it is true... But actually proving that, that is the problem..., I have to find out, to get someone to say actually "Yeah, that is the reason why I am doing this", which is quite difficult."

In response to the question, "What is an angle?" a Chief Sub-editor related the idea to the tragic events of the 11th of September.

> **QC** (by email) To start off with, there was the *straight* reporting of facts: a plane has hit the World Trade Center, then a second plane has hit the WTC... The *default angle* is 'what has happened'...
>
> ...But this soon develops, the new "*angle*" comes into play. I remember on the day that by the time of the second plane, I and others were saying: "*This must be an act of terrorism, because this is not coincidental, an accident...*". So had I been writing the story, I would have begun building up information to support my hypothesis that the acts of

> September 11 were terrorism. The standard journalistic questions of who, what, why, when, how would have been asked about the events against the backdrop of my hypothesis of terrorism.
>
> ... Essentially there is an *angle* to all news and features; it is really a working hypothesis that translates the gathered *facts*, which may include some speculation, into a coherent account.

An angle, then, can be a statement of fact, as QC describes it, "what has happened". Where it is less certain, it forms a hypothesis about what is believed to be the case; *i.e.* a refutable statement of fact—albeit one for investigation. QC described an angle as a hypothesis against the backdrop of which he would then be searching for information. In this sense, the angle acts as a macro-question supervening over, and prompting lower-level information needs.

The *determinant* of the constraint of following an angle is the angle itself. Deciding on an angle, though, is itself subject to some generic constraints. In particular, it must be *original*,

> **CJ** every news story has got to be new... and if it has been written before then maybe you can still write it, but you have to find a new angle to go in on.

true and *newsworthy*. AV paraphrased the newsdesk:

> **AV** ... "oh no, no, no, we want the truth, but we just want to make sure we get the right angle, I want a good punchy story as opposed to some boring rubbish." ... And at the end you could get it wrong in both ways ...

(these constraints, also apply to finished news reports and feature articles, and will be discussed in detail below).

Not only is following an angle a particularly important constraint, but the angle (*determinant*) can also be subject to mid-assignment change. There are a number of reasons for this. First, editorial decisions can be revised. DK said :

> **DK** The most likely change would be that the newspaper has decided to take a different angle on the story and that is, that

> tends to be the usual.. so the newsdesk tell me 'we want it done in this particular way or that particular way'. There are a number of times during the day when stories are discussed, but in particular, there is a morning conference and an afternoon conference. The decision might have been early in the day to approach a story from one particular angle, and by the end of the day, at the afternoon conference, which happens sort of between 4.00 and 5.00, there might be a decision to approach it in quite a different way...The conference is all of the editorial executives and the highest people, the editor chairs it, so.. Most likely, it seems to be most likely that something has grabbed them and the story about.. this is going to be a change of direction... quite often something like that will happen...

Second, through their research, the journalist will become deeply engaged with the story and they may find that the angle is wrong, or they may find information that suggests a better angle. CM said:

> **CM** Writing plans can change 'relatively frequently'. You might come up with what you think is a better idea than the one that the news desk gave you. You've then got to, of course, then got to go and convince the news-desk that you've got a better idea than the one they gave you. Which is not always easy. But you know, its something that should be encouraged. If you stumble across something that's more interesting than the original line that you've got then you've got to change it...

Third, a misunderstanding can occur between the editor and the reporter in the initial briefing. This misunderstanding is often only discovered once the story has been written and submitted. On this DK said:

> **DK** Any time I might be told 'the way you have done this is wrong'. And certainly, when I present my copy, it happens about once a week certainly. The News-desk themselves will just look at it and say that is the *wrong* way, that isn't an interesting way enough into the story, do it from a different angle that is more interesting.

Category 1.2 Deadline

Another important constraint is to meet a particular deadline:

> **JV** Because if you have got a big sheaf of printouts and you just need to refer to the stuff quickly and you just need to flick

through, and you know a lot of what I do is against deadlines, so you just need a quick reference.

LS You can't be inflexible as a reporter. And it should be in your job description, flexibility. It has to be, I mean there is no other way you can do the job, you have got such tight deadlines and you don't know what you are going to do from one day to the next.

Here the deadline is the *determinant* and the *source* is editorial decision making. A deadline for a feature article could be a few days, or for a news report would be a particular time on the same day. For news reports the deadline related to the page of the paper on which the report would be printed. Throughout the day, pages of the paper would be completed in the reverse order. The deadline for the last page could be around 3pm but for the front page it could be around 7.30pm. The available time could be as little as 20 minutes, depending on what time of day the reporter was able to begin writing.

Since the page on which a report will appear is decided on the basis of the perceived importance of the story, deadlines could be subject to some variation if, for example, events were to unfold in a way that increased the importance of a story.

Category 1.3 Word-count

An article should meet a particular word-count (*determinant*). Along with the angle and deadline, a reporter would be briefed initially to produce a certain number of words. Word-count decisions depended on the allocated column inches and were made by editors (source). The required word-count could be subject to change when layout decisions were revised.

Category 1.4 Prior written commitments

A basic rule of writing is that text must form a coherent whole. Hence, at any point during an assignment, written commitments can contribute to the shaping of future goals. One simple scenario that demonstrates this is where a sentence is started, but it is discovered that information must be sought before it can be finished. This was reported as an observation from the lab study in chapter 3 and was discussed by journalists during the current study.

Res Do you ever find that when... your halfway through when you realise there is some information that you haven't got that you need.

AF Yes.

Res Can you think of an occasion when that happened?

AF No, but there.. as you are writing the story, you will be writing a sentence and you think 'right, I don't actually know the answer to this' or 'I need to know a bit more about this.' So you go back to your contact, whoever it is that you have been speaking to, get it, add it in, and carry on.

Commitments, of course, can be undone (at differing levels of cost to the user), but ultimately some must be retained. They must contribute to a coherent whole, and they can only be made sequentially.

The *determinant* for the prior written commitments constraint is any previously written text. The *source* is the writer's previous actions. Clearly, change of the *determinant* over the course of an assignment is inevitable.

Category 1.5 Constraints on content

Category 1.5 draws together a number of lower-level constraints.

Category 1.5.1 Newsworthiness constraints

This category classifies some lower-level content constraints concerned with newsworthiness. (Some of the newsworthiness categories were developed in relation to the data using criteria of newsworthiness cited by Stephens (1993)).

Category 1.5.1.1 Timeliness/Currency constraint

The content of a report should be concerned with recent events.

RD ... they [colleagues] would be getting background information ... you could go down all the sorts of different corridors with the historical stuff, but the number one thing, the... obviously, getting the story to really cohere now, because the most important stuff would be the here and now.

AF If you can bring it on a bit, then fine, but once... it has been around however many years ago, you can't really do it.

Res It is just not new, it is not news.

AF No, it is history rather than news I suppose.

The timeliness/currency constraint expains why newswire *i.e.* short reports of breaking news are so important for triggering and shaping assignments:

HR: Breaking news items can come from the wires then the Newsdesk will assign a story or an aspect of a story - in the case of big stories - to individuals.

IC: When I get an assignment I will look at newswires to see what is being said about a story.

The *determinant* for this constraint is the time period since a given event, and the *sources* for finding them out are many. Since the time that an event occurred remains constant there is no indeterminacy.

Category 1.5.1.2 Proximity constraint

A report should prioritise issues local to the reader. Proximity here relates to physical location or more abstract feelings of group identity.

LS The main aim is to tell the reader what is happening. So first of all you say '2 people died, 7 people injured when 2 trains collilded'. And the next thing you have to say is how this affects people, the immediate effects...

...you know, I think we know, from what we know about our readers, what sort of things will interest them, and how you interest them and write the story. So, if works on that rather than... It is not a format, it is knowledge of what will appeal to the reader, what the reader needs to know, they want to know, would like to know.

Meeting this constraint may involve choosing one story over another or bringing particular issues forward in the reporting. The *determinants* for this constraint are the nature of the readership and the relative locality of the story. For the journalist, the *sources* are their model of their readers and the many sources they might have for finding out about a story. As the journalist's knowledge of the story evolves so might assessments of how well the story meets the proximity constraint.

Category 1.5.1.3 Exclusivity constraint

The exclusivity constraint prescribes that a report should be different from any piece published before. Interviewees said that they should not replicate any

story that had previously appeared in national newspapers or magazines; local and foreign newspapers were considered outside the scope of concern. This issue seemed to revolve around what their readers were likely to already know.

The question arose during data gathering of what it is that is compared—on what feature is the similarity judged? Clearly, a one word difference would not be enough, but another story touching on the same issues would not necessarily invalidate exclusivity. Although this question was difficult to determine, some evidence was provided by AF:

> **AF** ... if it has been written before then maybe you can still write it, but you have to find a new angle to go in on.

And also NH...

> **NH** It doesn't have to be.. every single thing doesn't have to be brand new, but it does have to have some new *twist* or *peg* or *hook* or *angle* that is different.

These extracts suggest that the feature of comparison is the *angle,* and so the exclusivity constraint was understood as prescribing that the angle of a report be different. Hence the *determinants* are the proposed angle for an assignment combined with the angles of previous articles in prominent national newspapers and magazines. The *source* is editorial decision making and the cuttings archive respectively. Causes of *a priori* indeterminacy are changes in editorial decisions and the fact that the journalist's knowledge of what has been published before can evolve during an assignment.

Category 1.5.1.4 Human interest constraint

This constraint prescribes that the content of a report should prioritise human interest. Like the proximity constraint, this requirement may be used to choose between stories or to bring out elements in an existing story. One way of doing this is by using anecdotes; as NH explained:

> **NH** ... anecdotes in particular in a newspaper, a little story, a little couple of paragraphs detailing somebody's experience, has great power to influence the reader. That is what journalism is all about. It is about raising emotion and getting people interested.

The *determinants* for this constraint are the human interest components of a story such as anecdotes. The *sources* for these can be many and depend on context. Indeterminability can arise out of the fact that a journalist's knowledge of a story can evolve throughout an assignment.

Category 1.5.2 Historical context constraint

A report should relate what is being reported to relevant historical context. This might include previously reported events surrounding the one being reported, or it may include previous events of a similar nature. As NH explained:

> **NH** How does it help you? It just puts the whole thing in context and enables you to interpret the latest story in the light of what's gone before. I mean, otherwise, you are constantly treating stories in a naïve fashion, and that is not the job of a specialist. A specialist should bring a depth of knowledge to the story and cast it in that context. But you don't always have that knowledge yourself, so you have to acquire it from somewhere else.

The *determinants* for this constraint are the story being reported and its historical context. *Sources* can be many and depend on the story but in many cases will be the cuttings archive. *A priori* indeterminability arises from the fact that a journalist's knowledge of relevant historical context can evolve throughout an assignment.

Category 1.5.3 Accuracy constraint

The need to report the truth accurately is a very important requirement. Even though the journalists interviewed agreed that newspapers can and do get it wrong, for their own part they were very concerned with accuracy. As DK said: "I have got to get all those facts right". AIF commented that this was particularly important when reporting legal proceedings. The need for accuracy also extends to getting name spellings right.

The accuracy constraint prescribes that, within a report, claims must be true, and name spellings must be correct. The *determinants* for these are the facts and correct spellings respectively, and the *sources* are various. A cause of *a priori* indeterminacy is the fact that a journalist's knowledge of the facts and name spelling evolves throughout an assignment.

Category 1.5.4 Legal constraints

A report should not transgress any laws; this includes the requirement of avoiding libel (category 1.5.4.1) and avoiding perjury (category 1.5.4.2.).

> **AP** There is a lot of information that couldn't go in the story because you would certainly get a writ because it is libellous. It may be true, but it is libellous and you can't do that.

> **LS** If you have got a running court story, you cannot add anything to it that hasn't been said in court in front of the jury, so all you do is colour your take on a story... I tend to avoid cuttings on court stories...

The *determinants* of these constraints are the respective *laws.* Since journalists are normally well versed in legal requirements these constraints are not normally a source of uncertainty and they do not change during an assignment.

Category 1.5.5 Explanation constraint

The explanation constraint prescribes that a report should attempt to *explain* events:

> **NH** ... we are always looking for explanations. If you come up with an explanation that might fit the present circumstances as well, that is good... people want explanations.

> **AP** ... so then I either do it myself or I get the library to... I want everything that is written about his house ... what people want to read is why it is so...

> **LS** But what you really want to be able to do is tell the reader exactly why this crash happened and what can be done about it to stop it happening again...

The *determinants* of the explanation constraint are the events themselves and the *sources* are various. Uncertainty in relation to this constraint arises from the fact that a journalist's understanding of why something has happened can evolve throughout an assignment.

Category 1. 6 Constraints on structure

Category 1.6 draws together constraints concerning report structure.

Category 1.6.1 Cut-from-bottom constraint

In a news report, information should appear in order of importance with the most important, new or unique information appearing first.

> **AF** I always start with an intro and then work downwards to the end and you know the best information goes at the top in a decreasing order.

> **PM** Well I mean, as I say it's discussed with the news desk what is the first thing to put in. What do you go in on. Go in on the fact... that is the extra bit of news so that is what you go in on, that is how you start.

One journalist described the order of information in a report being determined by a "pyramid of priorities". Sometimes the first piece of information can be the only new information to appear in the report, the remainder being composed of background in decreasing order of importance. In this respect, news reports were described as being relatively formulaic. Feature articles were described as conforming less to a prototypical structure.

Cut-from-bottom ordering was described by one journalist as partly a legacy of past printing technologies. Matching a story to a given area of page space was often a question of omitting the end of the piece, so this had to be the least important information. These days, reports can be cut electronically at any point, but the *cut-from-bottom* structure persists since it helps readers to gain the gist of a report without necessarily reading it through.

The *determinant* of the cut from bottom constraint, then, is the relative importance of different pieces of information and the *source* of this judgement is the journalist. The cause of *a priori* indeterminability is that throughout an assignment the journalist continually finds new information, and at any time some new piece of information can supersede another in terms of the journalist's estimation of importance.

Category 1.6.2 Original wording constraint

A news report or feature article should be worded originally. The *determinant* for this constraint is the wording of pieces published before and the source is the cuttings archive.

> **LS** I tend to print them out so that they are next to me... I very rarely cut and paste, partly because I am concerned about

> copyright, partly because I hate using other people's words when I can use my own words...
>
> **AN** I think that the information I take out of cuttings is this... I am not copying out chunks if that is what you mean. Some people do. When you read the cuttings you see the same bits copied out.

The cause of indeterminacy is that the journalist's knowledge of what has previously been written can evolve during an assignment. This constraint has become more important with the introduction of graphical user interfaces which allow the pasting of text from one document to another.

Category 1 pulls together constraints with diverse sources, all of which impinge on the journalist's activities. The list given here is not intended to be complete but has been chosen for the contribution it makes towards explaining the information behaviour that will be described subsequently. But despite its incompleteness, the list provides indication of the complexity of the journalist's task. The constraints present a landscape for the journalist to explore and negotiate. Knowing the constraints, and knowing how to address them is part of the professional knowledge of the journalist. By presenting the constraints not as known, static entities, but, as frequently only partially known and frequently dynamic, the account describes a part of the context for complex dynamic activity—the other part being the journalist's resources.

4.4.2.2 Category 2 Information Resources

Constraints have a key role in explaining task activity, but understanding why people do what they do, and crucially, why they do it the way they do it, depends on also understanding the resources that are available to them. The resources which arose from the analysis are all information resources of one kind or another (*i.e.* resources which allow the storage and retrieval of information).

In common with the Distributed Cognition approach (Hollan, Hutchins & Kirsh, 2000) the analysis attempts to soften the distinction between inside-the-head and outside-the head by considering both internal and external resources as a part of a single work-system. The first dimension, which provides the top-level decomposition, is *location* which specifies whether an information resource is within the journalist (*i.e.* cognitive, in the head) or outside. The next dimension,

mode of access, specifies whether the resource permits read-only access (*e.g.* a document) and read-write access (*e.g.* memory). The third dimension, *interaction paradigm*, is borrowed from a framework for information retrieval interaction by Bates (1986b). This dimension can have the values *passive* or *active,* where a *passive* interaction is where the user is prompted by the resource when salient information becomes available, and an *active* interaction is one in which the user must explicitly query the resource. The final dimension is *information scope*. This specifies the type and scope of information that is stored in the resource.

As with the constraints described above, the resources are not intended to be exhaustive, but rather provide an explanatory context for activities.

Category 2.1 External information resources

This category classifies information resources that are external to the journalist.

Category 2.1.1 External read-only information resources

External read-only resources are external to the journalist and, from the perspective of the journalist working on an assignment, limited to read-only access.

Category 2.1.1.1 Electronic news cuttings (ENC) service

An online cuttings archive is a database of past articles and reports from newspapers and magazines. It has read-only access and supports an active interaction paradigm using keyword/Boolean querying. The participants in this study had access to two ENC services: one in-house database containing articles from a small selection of key newspapers, and one commercial subscription service containing articles from local, national and international newspapers and magazines.

Category 2.1.1.2 News Library

As with most news media organisations, journalists at *The Times* had access to a news library. The news library would once have been the primary location for obtaining newspaper and magazine cuttings which would be stored in topic files for easy retrieval. When ENC services were first introduced, they would have been accessed in the library usually by trained expert intermediaries. Since the introduction of ENC services at the journalist's workstation, and

journalists consequently performing their own search and retrieval, both of these library functions had declined but not disappeared entirely.

> **VE** I would like to see how it has changed from the last time, and the way to do that is ask the library. Because I haven't got time to figure out all that, that is only one of god knows how many stories which might or may not be working on a particular day.
>
> **RB** I tend to use the library for searches that predate the Editorial Database and the internet, but not very often.

Value was placed on the library's collection of hard-copy files of hand-selected cuttings (which often predated the content of ENC archives and continued to be updated) and on the mediated ENC search service and search expertise that the library continued to provide.

Category 2.1.1.3 Newswire resources

This category classifies information resources based around newswires. Newswires are short, summary reports of breaking news stories which are provided via a feed by third party news agencies.

Category 2.1.1.3.1 Copy Taster alerts

A Copy Taster is a person in a newsroom whose job it is to track the assignments that different journalists are working on and to monitor incoming newswire feeds so that they can alert people when new information arrives relating to their assignment. Hence, from the perspective of the reporter, this is an external resource with read-only access that supports a passive interaction paradigm, the information scope of which is breaking news stories on a given topic.

Category 2.1.1.3.2 Newswire archive

> **SB** First off I will search news wires to see how much info I can get in before I start ringing people.

The newswire archive is a database of past newswires which can be queried by reporters at their desks. The database holds each newswire for about 30 days. This is an external resource providing read-only access through an active interaction paradigm.

Category 2.1.1.4 Informants

The *informants* category subsumes people who provide information to the reporter; this includes witnesses and experts and colleagues.

> **LS** A big story like that [train crash] (you've) absolutely got to get eyewitness reports.
>
> **VE** I would contact the embassy. But if I did that, again, I would seek the information from the expert who is closest to the subject matter.
>
> **Res** Do you ever find yourself looking for journalists who have written on a particular subject in the past?
>
> **AF** I would talk to colleagues on the paper and I wouldn't really go outside. .

They are external, read-only resources who support an active interaction paradigm. They differ in terms of information scope and tend to offer information about specific events from specific perspectives, detailed and sometimes technical information, and informed opinion on specific topics.

Category 2.1.2 External read-write information resources

This category classifies information resources which support read-write access. In other words, these are resources supporting information storage as well as retrieval. In all cases the interaction paradigm required is active. Their information scope is whatever the journalist has considered useful at some time during an assignment, and has stored.

Category 2.1.2.1 The 'holding document'

'Holding document' is an informal *in vivo* term. It refers to a word processor file used as a temporary storage space for information gathered from other read-only information resources.

The role of the holding document is to mitigate against limitations in the journalist's memory. The term 'holding document' was used by just one journalist, although other journalists used similar terms to refer to what was essentially the same concept; these included 'work-paste-pad', 'information basket' and 'kind of database'.

SB I would have two windows open and copy and paste wires and cuttings into a 'work paste pad' or whatever.

AA If there are just a few facts I want, I will just cut and paste and put it at the bottom of the file that I'm working in.

This external resource permits read-write access and supports an active interaction paradigm. Its scope is information considered potentially relevant at some time during the research and writing process.

Category 2.1.2.2 Printouts

As an alternative to a 'holding document', whole documents can be printed. They can be annotated and so they are classed as having a read-write access mode. Their scope is whatever documents and information were considered potentially useful at some time during the research and writing process.

Category 2.1.2.3 Written report structure plans

Many interviewees were asked whether they externalised a report structure plan. Although this was rare, a few did, and so it is included here.

CM I only write a plan for longer pieces. You get to learn what the right structure should be and you can hold this in your head.

LS I write one very occasionally when I am on the road, I jot down 4 or 5 points I want to make in order, but no, basically I don't do structure plans.

A written report structure plan is an external read-write resource supporting an active interaction paradigm with the information scope of report structure decisions.

Category 2.2 Internal information resources (knowledge)

This category classifies internal cognitive resources. The classification is made in terms of content (*e.g.* domain knowledge, writing knowledge), rather than in terms of human information processing theory classifications (*e.g.* working memory, long-term memory). All the internal information resources support read-write access.

Category 2.2.1 Domain knowledge

Domain knowledge is a journalist's knowledge of the world of current affairs, events of public interest and news in general. Depending on experience, it is always more or less partial and can be restricted to particular areas of interest.

> **BW** Yeah, I mean you always, you would have quite a lot in your head reading every day what everyone, all the nationals say about [225] issue, but also what the trade press is saying, but you know we'd always do a cuts check.

Category 2.2.2 General writing knowledge

General writing knowledge is a (more or less partial) knowledge of conventions and methods that cover writing in general.

Category 2.2.3 Newspaper writing knowledge

Newspaper writing knowledge represents a (more or less partial) knowledge of writing conventions specific to writing for a newspaper. This will include, for example, knowledge of many of the constraints listed above such as the *newsworthiness constraints* and structural requirements such as *cut-from-bottom*.

Category 2.2.4 Resource knowledge

Resource knowledge refers to a (more or less partial) knowledge of the methods required to operate tools such as the available technological resources.

> **NH** It is true of all the systems. You get to a certain point of competence with them and there is no real advantage in getting any more competent if what you are learning is something you are only going to use once a month. If by the time you come to use it again, you will have forgotten. So all you learn is the things you used every day, or at best every couple of days...

Category 2.2.5 Internal report plans

As discussed above, the journalists interviewed said that they rarely if ever wrote plans for their reports Nevertheless, as they sought and gathered information their mental concept of the content they would write and how it would be structured would evolve.

> **JG** I don't write a plan. The hard bit is discovering the information and once you've got that the story just 'writes itself'.
>
> **Res** Can you say why you don't do that? [write a plan]
>
> **AF** Well mostly it is just training and common sense, depending on the story. You know what you have to do to get the information you need... You just automatically know what you have to do for each scenario that you get, because it is not always the same thing that you are doing, so it is not 'right', it is not every story would be step one, two, three, four and five. One story might need step five and six, on another story you might need step one and four.

The constraints and resources combined are the context within which a news reporting or feature writing assignment takes place. They provide the context, motivation and boundary conditions for a process, and hence provide a way of understanding and explaining it. In a sense, they represent the task situation, or rather, the task itself. Information behaviours are designed to address that task situation by bringing about change.

4.4.2.3 Category 3 Information Behaviours

Category 3 classifies information behaviours that were identified during the study into three major groups, these being emergent from the analysis. The groups are: *Information-seeking*, *Information-gathering* and *Information Reviewing*. Information-seeking is concerned with finding information. *Information-gathering*, which corresponds with Bates' Berrypicking (Bates, 1989), is concerned with collecting information from a source location that is to be retained for a current assignment, and information reviewing is concerned with relocating information that has already been encountered (both gathered and non-gathered).

The only common dimension for all the behaviour categories is *Goal*, which has the value *finding information*, *gathering information* or *relocating information* depending on the top level category. All the categories have a common set of properties (axial relations) based on the constraints and resources framework. The axial codes (with short descriptions) are shown in table 4.1.

The classification differentiates information behaviours; implicit in this is the idea that each can, in some sense, be considered independently. But this is not to

say that any observed behaviour trajectory will necessarily correspond to one, and one only information behaviour—this is not the kind of independence that is implied. Rather, activities are seen as abstractions with any given instance potentially relating to a number of behaviours. The differentiation implied is *logical*, *i.e.* that one information behaviour does not *necessarily* entail another. By this is meant that each is understood as having the *potential* to be realised independently, but, in practice, this need not be the case.

Axial code	***Description***
Rationale:	The reasons for conducting an information behaviour.
Circumstances:	Details the circumstances under which an information behaviour is conducted. This is closely related to rationale offering further qualification of the preconditions.
Means:	Details the way that the information behaviour is performed, including resources and specific strategies with those resources.
Consequence:	Lists any consequences of the information behaviour over and above the rationale *i.e.* unintended consequences.

Table 4.1 The axial coding framework
used to code information behaviour categories

In this sense any observed behaviour might be classifiable under more than one category. For example, an information behaviour might address one goal and use another or a series of others as its means. Like Ellis' (1989) analysis of social scientist's information-seeking characteristics, it is not intended that the model should be taken as stating any particular sequence or sub-goal embedding.

A single behaviour may also address multiple goals. This is a phenomenon known as 'polymotivation' (Cole, 1996). From the interview data, polymotivation featured within many of the described information retrieval activities. Indeed, given the user costs of creating searches and reviewing results lists, the integration of multiple goals within a single broad trajectory of search behaviour may represent a good economy of action.

Category 3.1 Information-seeking

Information-seeking behaviours are characterised by the goal of finding information. Dimensions used to differentiate them are *focus*, corresponding to how focussed the need is (low, medium or high) and the *extent of searching* (light or thorough).

Category 3.1.1 Exclusivity checking

Motivated by the exclusivity constraint (constraint category 1.5.1.3), *exclusivity checking* was described with few exceptions as the first thing that a journalist would do on beginning a new assignment. One journalist described it this way:

> **AP** Well first I like to check that nobody has beaten me to the story, because that would look very silly. That doesn't always stop you doing it, you know if it is buried on page 19 of the Independent, nobody will remember. Some papers don't use stories properly, they don't give them the prominence they deserve.

Occasionally exclusivity might be traded-off against other newsworthiness constraints, although discovering that a report had previously appeared would usually lead to the idea being abandoned or modified. CJ explained:

> **CJ** So the first thing that I will do is to check on the database past stories to see what has already appeared. One of the reasons for that is to make sure that whatever you are writing hasn't already been written before...

CJ indicates that originality checking is only *one* reason for this initial search. In many of the accounts of search information behaviour, interviewees described pursuing multiple concurrent information goals (polymotivation). In the following extract, MG describes extending this initial search motivated by the goal of developing a better personal understanding of an issue:

> **MG** Obviously the main interest is whether it has been in a British newspaper. ...but I like to know whether it has been in the LA Times and the cutting might well tell you something useful anyway... it might give you background on the stories... things in the background that are not apparent to you when you are looking at the thing to write a story...

And in the following extract, DI describes integrating originality checking with the gathering of potential content that she might later include in her copy:

> **DI** ... the first thing you do is go into your database... to find out if a similar story's been written before... and erm... just to see maybe if another story's touched on it in the past, say, that you can pull out bits from that and add it to your story.

Circumstances where exclusivity checking was not performed were those where exclusivity could be assumed. For example, if the journalist was working on a piece based on a research article in an academic journal, it might be assumed that the journal's review process would ensure novelty. Also, there was a tendency for experienced, specialist journalists, with good domain knowledge (resource category 2.2.1) to feel more confident about making exclusivity judgments without checking.

In the discussion of the exclusivity constraint (above), it was shown that the characteristic of a story which is required to be original is the angle (constraint category 1.1). A report or feature may well, and often will, repeat information which has been reported elsewhere. However, the angle, or central thrust of the story, must be new. Consequently, exclusivity checking involved researching the angles of previously published reports within a specified topic range. The means used for checking the originality of an angle was predominantly to search one of the ENC archives (resource category 2.1.1.3.2) at the journalist's workstations, or to delegate the job to a librarian (resource category 2.1.1.2) who would do likewise. Given that the angle is normally expressed within the first sentence of a news report, ENC search results lists which provided the first sentence of each article as part of the document summary were described as being particularly convenient for scanning the angles of many reports quickly and were described as saving considerable time.

Category 3.1.2 Background information-seeking

In addition to exclusivity checking, seeking background information on a story was described almost universally. The principal resource for finding background information on current news stories is the ENC service (resource category 2.1.1.3.2). This is how BW responded to a question about background searching ("cuts check") using the ENC service:

> **Res** So once... the story gets started. Do you do background searching?
>
> **BW** Yeah, I mean you always, you would have quite a lot in your head reading every day what everyone, all the nationals say about an issue, also what the trade press is saying, but you know we'd always do a cuts check.

Part of the value of background information is to provide deeper knowledge of the context of an issue (resource category 2.2.1) to enable better interpretation of recent events (constraint category 1.5.2). In the following, NH relates the understanding of background information to providing depth to a story:

> **NH** How does it [background information] help you? It just puts the whole thing in context and enables you to interpret the latest story in the light of what's gone before. I mean, otherwise, you are constantly treating stories in a naïve fashion, and that is not the job of a specialist. A specialist should bring a depth of knowledge to the story and cast it in that context. But you don't always have that knowledge yourself, so you have to acquire it from somewhere else.

It was explained that some background information from ENC archives would, in nearly all cases, be "woven" into a report.

Another commonly cited reason for background information-seeking was as preparation for conducting an interview—a valued but short information-seeking opportunity in itself. Here, raising domain knowledge (resource category 2.2.1) was again the issue. This, in addition to when a story was controversial or technical or when the longer deadlines (constraint category 1.2) of feature writing permitted, provided the conditions for thorough and extensive background information-seeking. One science correspondent said of her interview preparation on a technical subject:

> **GQ** I certainly wouldn't like to have spoken to him without having researched the subject before, because I didn't know anything about it and I wouldn't have known the questions to ask.

Another interviewee said that before interviewing she would want to "have found out every cough and spit in advance... so you can frame your questions cleverly".

Most frequently, a good deal of time would be spent on background information-seeking, although time constraints (constraint category 1.2) would often be restrictive. When writing on less serious topics, and where interviews were not planned, background information-seeking could be more modest with

the journalist skimming a few documents and "pulling out odd bits of information".

> **NH** I would say you are usually skimming for interesting facts. You are usually looking for a few basic facts and the odd anecdote or example to illustrate a point, so you skim the [unint] of it for those things and you just ignore everything else.

Most commonly, journalists said that they would search ENC archives (resource category 2.1.1.3.2) themselves, although some delegated this task to the library (resource category 2.1.1.2). Also, one experienced specialist who had accrued many contacts said that she would tend to contact an expert (resource category 2.1.1.4), such as someone from an appropriate government organisation, who could provide her with a "background briefing".

Category 3.1.2.1 Seeking background overviews

In the discussion of the field study in chapter 3 some questions were raised about the need for overview documents. These questions were: Do journalists have these needs? If so, how are they resolved? And how might they be optimally resolved? The first two of these questions are addressed.

Background overview documents were considered useful by the interviewees, particularly at the beginning of assignments and for journalists with lower domain knowledge (resource category 2.2.1), such as novices, generalists and specialists working "off-patch".

> **Res** Would you look for a background overview in circumstances other than if you were going to do an interview?
>
> **VE** Well if I was suddenly sent out to cover the Paris office again or some other part of the world which I am not familiar with, then obviously I would ask the library to do me cuttings of all the things that have happened in the last year or something, so I knew the personnel, I would contact the embassy, I would contact the, you know...

To some extent, the library service (resource category 2.1.1.2) had anticipated this need by compiling a small collection of "fact files" each providing a basic outline of a "hot button" subject. Compiling these files, however, was labour intensive.

It was often felt that the best "way in" to a subject at the beginning of an assignment was to read one or more documents providing an overview of the topic, but finding them was described as "difficult".

> **AA** Sometimes you want an overview first, like a big piece on the subject.

The available ENC services provided poor support for finding overview documents since it was not possible to create searches that would discriminate them automatically. Consequently, the journalist would have to search an ENC service for all documents on a subject, and then browse the inevitably extensive results list. In the following extract, a journalist who had been working in Britain for only a month describes this process and explains the sort of documents he found helpful:

> **CM** Well considering I started with zero knowledge erm... that's what I had to do [find an overview] in a pretty quick situation.... And there may have been a couple of hundred stories at least there. Where you see the first line, you can get a bit of a view about what sort of a story they are. And you look for something that has perhaps been in *The Times, Telegraph* or a Sunday edition which is a bit more of a backgrounder erm... and... rather than going through every single story I quickly found two or three decent feature length pieces which gave me a pretty good background...

Notably, CM was interested in an article that was more of a *backgrounder*[1], and expressed some confidence in being able to assess relevance based on the first line of the text. He also expressed an interest in finding articles of a reasonable length, and said that his need was ultimately met by two or three feature length pieces.

In the next extract, NH expresses a preference for overviews in broadsheets.

> **NH** You know, if somebody has done a lot of research and put it in the FT or somewhere it is jolly helpful, because otherwise you have got to try and do it on the phone and you

[1] A *backgrounder* is a piece which supports a main *lead* article in a newspaper by providing background information, often appearing in a sidebar beside the lead article.

> haven't got time to do that, so you use other people's research if you think it is sound.

He expanded on this by saying that he would also try to focus on articles by known specialist journalists. Other ways of obtaining a background overview included looking at recent newswires (resource category 2.1.1.3.2) on a subject and talking with informed colleagues (resource category 2.1.1.4). Occasionally, a newspaper article would provide an explicit chronology of events and these were valued.

Often, a journalist would seek an overview piece specifically about a person, such as a profile, biography or the last few big interviews.

> **NW** ... you know I was looking up stuff on Passport to Pimlico, the old sort of Ealing comedy the other day. I was going to interview Susan George, so she is a bit before my time, so I was ... something Straw Dogs was vaguely in the back of my memory so I needed to check her biography and stuff and find out what the films should be.

This was particularly important when the journalist was writing a profile or obituary themselves. Similarly, they would often search specifically for profiles of specific companies or organisations. ENC service support was described as poor for finding profiles. Strategies that were described included using the person's name in the query in addition to the term "profile", but it was acknowledged that this tended to result in a low recall search.

Category 3.1.3 Seeking evidence for a hypothesis

Background information helps the journalist provide depth in their reports and helps them to formulate questions for interviews; but by definition, it is peripheral to the central point communicated by a report or feature. Information from cuttings can take a more central role where the central statement of an article is in itself part historical, such as the identification of a pattern or trend. In these cases information-seeking using cuttings archives (resource category 2.2.1) might be performed in order to identify evidence for a hypothesis, and often this will be the proposed angle of the piece (constraint category 1.1).

The journalists described seeking evidence for an angle as a frequent behaviour. They often referred to looking for 'facts and figures' to support a perspective. AT explained:

> **AT** ...it is not uncommon you are told by editor, you know, ten children have killed other children in the past fifteen years, go and find out more about that. So you are looking for facts and figures. It may not turn out to be true, and in the end the facts may not support the chosen angle, so then you have to change the angle.

And DK said:

> **DK** Well usually it will be an idea from an executive on the paper, to look at a particular area that is in some way scandalous or interesting, and then my job is to look into it to see whether their ideas are right, or if they have got the right end of the stick, and to try and find enough facts and evidence to backup their idea.

As NH said above, seeking evidence for an angle often involves *"skimming"* or *"sifting"*:

> **NH** I would say you are usually *skimming* for interesting facts. You are usually looking for a few basic facts and the odd anecdote or example to illustrate a point.

Searching for information to support a chosen *angle,* whilst being more focussed than background searching, was nevertheless somewhat exploratory in that it depended upon the recognition of valuable information as it was encountered.

In the discussion of constraints, the angle was described as often being subject to mid-assignment change. Two explanations for this were that the journalist, through their research, might find that an angle is not true, compromising the accuracy constraint (constraint category 1.5.3), or they might find information that suggests a better angle, creating an opportunity for better optimising newsworthiness constraints (constraint category 1.5.1). This process is represented in figure 4.2 as a cycle.

In the cycle the angle (a constraint) motivates information-seeking (an information behaviour) which develops the journalist's domain knowledge (a

resource). But this can result in new judgements about what the angle should or shouldn't be (*i.e.* changing a constraint). Hence a constraint is revised as a result of the information behaviour it motivated.

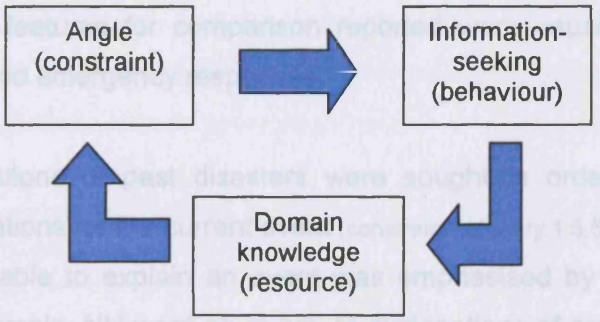


Figure 4.2 The angle drives information-seeking which develops domain knowledge, sometimes leading to a revision of the original angle.

Category 3.1.4 Information-seeking for feature comparison

In *information-seeking for feature comparison* the aim is to find similarities and/or differences between two events of a similar type on some shared variable or dimension. This may be performed as a part of background information-seeking or seeking evidence for an argument.

For these activities the search targets tended to be specific (facts and figures) and so resulted in a more highly focussed search than the types of background searching and seeking evidence for an argument described previously.

Category 3.1.4.1 Seeking properties of past disasters

Res When there is some kind of medical disaster in the world, do you find yourself looking for past similar disasters?

NH Yes, we would certainly do that. Say what happened last time, that is a very good way of doing it, say like foot and mouth, we were all looking back to '67 and the last foot and mouth epidemic, how was that handled, how was it different. Yes you do look at similar things from the past, if they have been similar.

Disaster reporting is a frequent job for a journalist on a national newspaper and the need to relate these to previous disasters of the same type (constraint category 1.5.2) results in some information needs based on specific properties for comparison. Sometimes the aim is to see in what way a recent event might be remarkable and so the feature of comparison is not specifiable, but sometimes the participants were able to be fairly specific about the features of interest. Specific features for comparison reported were causes, locations, casualty figures and emergency responses.

Causes and locations of past disasters were sought in order to suggest candidate explanations for the current event (constraint category 1.5.5). The priority placed on being able to explain an event was emphasised by a number of reporters. For example, NH went on to say of explanations of previous similar events:

> **NH** Well, it is nice to know them because we are always looking for explanations. If you come up with an explanation that might fit the present circumstances as well, that is good... people want explanations. They want to know why—why is there an outbreak of foot and mouth? why are we... you know why are we getting CJD? There is not much point in writing a story if you don't at least seek an explanation. There might not be one, in which case you have to say that there isn't at moment, but scientists are working in this way or that way to try and find one. If there is an explanation, it is jolly nice to have it.

If an explanation is found for a previous similar event then this might indicate liability due to the lack of a previous strategic response. BW explained:

> **BW** ... if it is a signal passed at danger or SPAD as it is known in the business, I want to know the history of that signal, so I will look for any other incidents in that area. For instance, if it is in Paddington, there are 67 incidents of signals being passed at danger in that stretch of track, a very complicated stretch of track, in the six years before the Paddington train crash. That is relevant. If you can find enough cuts to tell you that. So you look for, you know, previous similar causes.

Comparing casualty figures and the level of emergency response were also described as important since these are used as a broad metric of disaster severity and help to put a current event in context (constraint category 1.5.2). LS said:

> **LS** ... then you start looking at things like... 'is this the biggest crash ever?' which is like record breaking features about it, like was it the most deaths in Britain, the most injured, anything like that.

Category 3.1.4.2 Discovering/confirming what someone said

The exploratory study reported in chapter 3 suggested that quotation-seeking might be a common goal during news report and feature article writing. Whether this is so is a question that was left for the current study to address. It was found that seeking quotations was indeed a common goal.

> **Res** Would you want to find a quotation in a cutting? Is that something that happens very often?
>
> **AT** Yes.
>
> **Res** When would that happen.
>
> **AT** Well people say, for instance, Robert Winston has said the NHS is *terrible*, right, in this famous quote. So you really need, if at all possible, you need to find the quote. I mean sometimes they didn't really say what people remembered them saying. That is quite common.

In the extract above, a feature writer said that she often wants to find a quotation in cuttings. In this case, she has an idea of what was said, but needs to be sure (constraint category 1.5.3). AT's goal, then, is one of confirmation—comparing what was said to what was believed to have been said. This type of scenario was reported commonly, as was the scenario of comparing something said recently with something said in the past; this was particularly common with political stories (constraint category 1.5.2.). Indeed, at the time of the study, the paper was preparing for a national election campaign by manually creating a resource for cross referencing politicians statements with what they might have said in the past on the same subject. SC explained:

> **SC** I am going to be going down to the *'election bunker'* and much of that work will be compiling quotes and cross-referencing them. You know - "What did he say in 1997? And how a line's slightly altering." There's that kind of verbal trickery. The only problem is that some quotes 'hang around' and get repeated and it's difficult to pinpoint when it was first said.

Category 3.1.5 Confirming names and how to spell them

The need to locate proper name spellings, identified in the exploratory study, was recognised as a potentially problematic goal. This raised the question for the field study of whether this is indeed something that journalists frequently need to do. And if so, how do they do it?

In the field study, the need to locate proper name spellings was recognised as a frequent information behaviour. For example, NW, a freelance reporter, said:

> **Res** Checking name spellings, is that something that you find yourself doing?
>
> **LV** Yeah, all the time... I would hate to send a piece in that I feel I haven't checked the name thoroughly... In any field of journalism that I was working on I would check and double check all names, I would like to think... It is just professional.

Given cultural differences in how phonemes are spelt, the absence of international names from standard spelling checkers and English dictionaries, and the possibility that a journalist on a national newspaper may find themselves reporting foreign news, it is unsurprising that proper names are singled out as presenting a particular problem from the point of view of spelling.

LV's strategy for confirming name spelling using an ENC service (resource 2.1.1.1) was to perform one search for each variation he considered plausible. He would then use the reported number of hits for each query to assess which variation he considered most likely. Others used the same strategy using the Web. It was recognised that this strategy was fallible and, depending on comparative hit rates, could often provide ambiguous results.

Category 3.1.6 Identifying useful contacts

For a journalist, a contact (resource category 2.1.1.4) is a highly prized and often jealously guarded resource and would often be sought or opportunistically identified during background research. Some discussed the value of finding the names of people who had been involved in crime cases such as investigators or victims, others discussed finding the names of other journalists who had written knowledgeably about an issue. A particularly common

information behaviour was to find the name of an expert who could provide informed, balanced comment on a complex, technical issue.

Many strategies were used to identify useful contacts including asking colleagues, contacting press offices, contacting known experts in related fields for referrals or searching websites. Some interviewees reported using ENC services (resource category 2.1.1.1) to find experts by searching for other expert comments on the same subject and assessing the quality of the comment; if the comment appeared good then they would contact them on the new issues.

> **JV** ...that is another example, is looking for names of people who have commented on a particular subject in the past....if I am doing say a background on a big crime case, you know, and I am spending quite a lot of time researching it, then I would look at all the cuttings and look for people that I could target to ring up to go out with or to actually meet face-to-face to develop.

Other interviewees reported opportunistically identifying contacts while reading news cuttings.

Category 3.2 Information-gathering

Information-gathering is an information behaviour which integrates closely into information-seeking. In the discussion of the exploratory study this was identified as an issue for enquiry in the field study. The questions that were raised in that discussion to be applied to the field study were: How do journalists transfer information into a task specific repository, and what form does this repository take? This category answers these questions.

Information which had been found through information-seeking, and identified as potentially useful would not necessarily be used immediately. The interviewees reported that as information-seeking progressed selected information would be gathered and stored for later use. One journalist even used a metaphor almost identical to Bates 'Berrypicking' (Bates, 1989a):

> **RB** For historical pieces, I would see what cuts come up with and 'cherry pick' the good bits.

Where the journalists were seeking information from cuttings using ENC services (resource category 2.1.1.1), information would be gathered either by

dragging-and-dropping text extracts from a source document viewed within a browser into a 'holding document' (resource category 2.1.2.1.1) or by printing out the document (resource category 2.1.2.1.2) and highlighting selected key extracts (These methods form the subordinate categories for *Information-gathering*).

In each case, the goal of information-gathering was to make selected information easily available at some later time. In particular, interviewees described the need to be reminded of particular facts and figures (which are easy to forget) such as ages, details of jobs and careers, locations, so that these might be 'woven' into a report. Good quotations were also selected as were chronologies.

The ability to 'pick berries' successfully depends upon the ability to judge good berries from bad. Likewise, information-gathering depends on the ability judge the value of information as it is encountered; such judgments, however, can be hard to make. First, and as described above, an assignment angle (constraint 1.1) which "translates the gathered *facts*, which may include some speculation, into a coherent account" can change as a result of editorial review motivated either by the editors or the journalist. Second, the angle, as an initial conjecture, stands only as an outline of the finished report, and despite the fairly formulaic structure of news reports (constraint category 1.6), as the journalist works on an assignment, their concept of which details to include is constantly shaped and reshaped. As CJ explained:

> **CJ** ...the ideas will take shape all the time... at the point that they change all the time. It is only really when you have to sit down and actually write it that I would have to decide what way to go into the story... I am preparing it... I am preparing all the time.

As a consequence, information is usually only gathered on a relatively provisional basis. Not all of it will be used—rather, the gathered information is more of a series of potential use options.

Category 3.2.1 Dragging and dropping

When searching for information on ENC archives (resource category 2.1.1.1), source documents would initially be viewed by the user using a browser. A common method for gathering information from documents viewed in this way

was to use the Windows system 'clipboard' to drag text extracts from the browser into the 'holding document' (resource category 2.1.2.1.1). This was preferred when time was short (constraint category 1.2) and where only a small amount of information from a source document was considered useful, and also where there was no requirement for the journalist to be mobile (for example, leaving the office to perform an interview or attend a press conference).

> **AA** Sometimes I will print stuff out if there is a lot of information in the article that is useful. This is because I will be switching backwards and forward from the source to her piece. If there's just a few facts I want, though, then I will just cut and paste and put it at the bottom of the file that I'm working in.

> **JV** Yeah, if for instance I have been briefed to go on a story the day after, for instance like the High Court or something like that where I am out of the office, I might take a print, a few printouts with me, you know on the subject. So yeah I will take printouts with me to a job outside, but if I am in the office and I just need to cut and paste I just highlight the stuff I need and put it into a big file.

In order to support frequent dragging and dropping from a browser window into a word-processor ('holding document') window with ease, some of the journalists interviewed described sizing and arranging their browser and word-processor windows into a vertical split screen arrangement. Others, whilst recognising the desirability of such an arrangement, lacked confidence in their knowledge of how to set their screen up in this way (resource category 2.2.4).

Category 3.2.2 Printing

Printing documents and then highlighting the paper printout (resource category 2.1.2.1.2) in order to draw attention to particular extracts tended to be used when more time (constraint category 1.2) was available, when a large amount of information in a document was considered useful, or when there was a mobility requirement. Reading from printouts was generally considered more comfortable and print-outs were also seen as a better option for supporting the possible need to substantiate sources later in case of a subsequent challenge from an editor relating to accuracy (constraint category 1.5.3), a complaint from a reader, or a legal challenge (constraint category 1.5.4). Printouts with extracts

highlighted were in some cases reported as supporting greater ease of reference than computer based text files. DI said:

> **DI** I usually print cuttings because you can highlight them and it's quicker to have the page in front of you rather than to scroll on screen. If you didn't print them then you would have to go back and do the search again to find the info.

One journalist even reported reading a document in paper form and, on finding information he wanted to gather, dragging and dropping from a screen version.

Category 3.3 Information reviewing

Information reviewing is a category which classifies activities having as their goal the relocation of information already read. These are: *relocating gathered information*, *relocating read but not gathered information* and *reviewing for omissions*.

Category 3.3.1 Reviewing information gathered during an assignment

Information-gathering is performed so that information can be relocated later more easily than it would have been if it had been left in its source location. Where information had been collected in a 'holding document' (resource category 2.1.2.1.1), the journalists would often work with their screen split vertically between this and their developing copy file as they were writing. In this way they would be able to quickly review and be prompted by extracts and 'weave' pieces of information into their developing copy. Where cuttings had been printed (resource category 2.1.2.1.2) (and highlighted) they would be arranged near to the computer screen so they could be reviewed easily.

> **LS** I tend to print them out so that they are next to me.

At times during writing, cross referencing what was being written with what had been read and gathered was frequent and rapid.

Some interviewees reported that during the latter stages of writing they would review all their gathered information to see if anything had been omitted which should be included. In this case the 'holding document' and any highlighted extracts in gathered printouts collectively acted as a checklist of potential content. EJ explained:

> **DI** I go through documents highlighting important bits and then I go back through as I'm writing. I go back through as I'm writing to make sure I have included everything-as a checklist

Category 3.3.2 Reviewing information read but not gathered during an assignment

The need to relocate previously seen documents was seen in the exploratory study. Although the subjects in that study were provided with the means for recording information as it was encountered, when they came to write, both wanted to review source documents. This raised questions about the more general applicability of document review in journalistic writing. The questions in the discussion of chapter 3 which relate to the field study were: Do journalists find that they need to refer back to previously seen source documents? If so, why does this happen–to what extent is it due to *extraction omissions, new information goals* and *extraction errors*? Answers to these questions re-emerged in the field study in this section.

Potential failure to make the 'right' relevance judgment when information is encountered was described above as arising for two reasons. On the one hand, uncertainty and change with respect to the pursued angle (constraint category 1.1), and on the other hand the fact that the journalists plan for their report (resource category 2.2.5.2) is constantly being shaped and reshaped. Both of these mean that journalists do indeed need to review documents they read during an assignment in order to find more or different information. DI said:

> **DI** This [angle change] means that sometimes you have to do some more information searching. This can mean going back through your cuttings to see if there's anything else you want. The information that you already have can reduce in importance.

A testament to the report plan as an evolving, developing idea is the fact some participants reported reading a cutting and failing to collect information that they would later consider important and need to find again. This might be described in terms of failing to see the significance of information through lack of a full grasp of the subject. RG said:

> **RG** When I was on the news desk, you might stumble across something on the web that is the absolute bottom line on a

> subject, but you don't realise this because you don't have the full grasp of it all. But as your understanding forms you suddenly realise the value of the previous source. And then you're struggling to get back to it.

One news reporter also described consistently failing to record the date of an article, and a Features Writer said:

> **AA** Sometimes you remember seeing a fact that you forgot to copy and so you have to go back in (to an electronic cuttings archive) to find the article it was in.

NW, the freelance reporter explained that his notes (resource category 2.1.2.1.1 would often be 'sketchy'—that he would often use the information that he had gathered as "signposts" more than anything (partly because of the large amount of information that he would be confronted with during his research and his inability to judge relevance early-on) and that he would often need to refer back to source documents.

Category 3.3.3 Reviewing information read prior to an assignment

On occasion, a journalist may recall a news or feature article which they had read some time before an assignment which they later recognised as useful for that assignment. They would then want to review the article in the light of their new task and would probably use an ENC service (resource category 2.1.1.1) in an attempt to retrieve it. For example, a science feature writer said:

> **AA** ...there was an instance of this the other day, actually, when I know I'd read something on paper and I wanted that fact. Now I couldn't remember exactly what it was but it was relevant to what I was doing and I wanted to include it, and I couldn't find it; I just couldn't find it. Er... you read so much during the course of the week. I couldn't remember whether it was an American newspaper or a British newspaper, a magazine or what.

4.5 Summary and discussion

At the start of this chapter its aims were set out as that of confirming information behaviours identified in the exploratory study, of discovering others, and of seeking to explain them in terms of the journalist's task situation (research

question 1). The chapter also set out to consider how the knowledge gained relates and contributes to more generalisable information behaviour theory (research question 2).

The information behaviours raised by the exploratory study for *in vivo* corroboration and development were: *biography seeking* (which was generalised to *seeking an overview*), *quotation seeking*, *confirming proper name spellings* and *information-gathering.* These behaviours have indeed been confirmed and developed in the field study. The analysis of the field study data identifies many information behaviours and the analysis has sought to explain them in terms of the constraints and resources which together make up the journalist's work context.

4.5.1 A model representing dynamic interaction between behaviour, constraints and resources

In figure 4.3 the findings are summarised and represented graphically in the form of a model. The model represents the main elements of the findings, constraints, resources and behaviours, as parts within a system of interacting factors. The identification of constraints and resources, and the recognition of these as dynamic, provides the key to understanding the system as a dynamic and changing whole.

Behaviours are shown at the centre of the model embedded within the context of reporting constraints and information resources. The behaviours are classified into their three major classes: information-seeking, information-gathering, and information reviewing. In the figure, behaviours cascade from left to right to suggest a sense of typical temporality (*exclusivity checking* is usually the first thing done with *reviewing for omissions* usually last). However, the temporal structure is intentionally vague in order to accommodate flexibility and adaptation on the part of the researcher/writer in deciding what to do and when. Also, it is stressed that the information behaviours should not necessarily be considered as mutually exclusive. For example, *information-seeking for feature comparison* may be done in order to provide evidence for an angle.

The constraints (shown at the top of the model) include many factors originating from many sources and indicate the complexity of the journalist's task. The

constraints range from the relatively local and transitory, such as prior written commitments, to the relatively global and stable, such as legal requirements. The information resources (shown at the bottom of the model) can be external to the individual, such as an ENC archive and informants, or internal such as domain knowledge and general writing knowledge. They can support active interaction, such as the library, or passive interaction, such as the copy taster alerts (a sub-type of newswire resources).

Although the journalist typically understands the constraints of news and feature writing when these are considered in general terms, knowing the implications of these requirements at any one time during any one assignment may be unclear; this is understood in terms of the *determinant* for each constraint being more or less certain. A preferred angle may not have been communicated by an editor, or the truth against which accuracy is measured may not be completely known. Hence, the task is uncertain and this uncertainty inevitably causes uncertainty in the information behaviour.

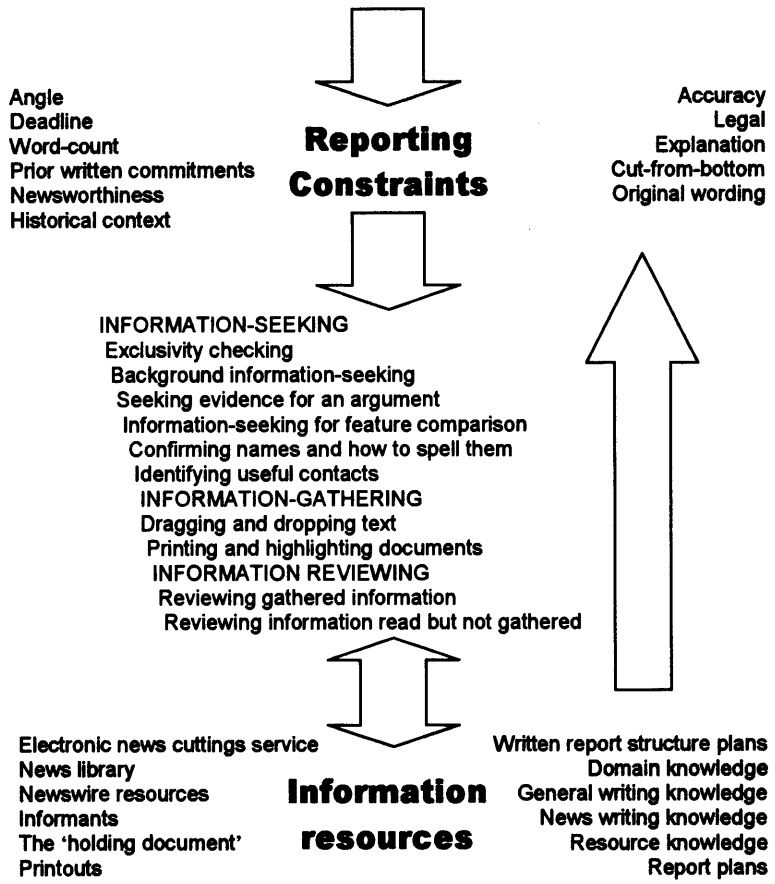


Figure 4.3 A summary model of journalists information behaviour influencing and influenced by dynamic constraints and resources

Closely related to this uncertainty is change. Through the central behaviours, uncertainty is resolved as the task evolves. Also, external influences can change the task. The model generalises dynamic influences into five types. These are represented by the arrows between element groups. The types of influence are:

External influences on constraints

The white arrow pointing down to the reporting constraints represents the fact that the determinants of many (or most) of the constraints have external sources. Further, part of the dynamic and often uncertain nature of journalistic research and writing arises from the fact that these external sources can influence the constraints mid-assignment. For example, the journalist's work is a component within a wider process newspaper production, and external events in the wider process can change the nature of that work; editors can change their minds.

Influences of constraints on behaviour

The downward arrow between constraints and behaviour represents influence propagating down from the constraints to motivate and structure the journalist's behaviour. This is the influence which is implicit in the understanding of a factor as a constraint. The constraints collectively define a complex criterion by delimiting the space of possible solutions into more or less satisfactory possibilities. A constraint may lead to information-seeking activity, for example, by determining a required angle or word-count, or by determining the requirement of locating a story within a historical context.

Influences of resources on behaviour

The upward arrow linking resources with behaviours represents the influence that resources have in determining how activities are performed, and, given user-judgements of cost and benefit, the extent to which they are performed, or even whether they are performed at all. Resources provide the tools with which behaviours are performed and, through their characteristics, delimit the possibilities of what can be achieved.

Influences of behaviour on resources

Whilst resources delimit and determine the cost of what *can* be done, behaviours can influence this by developing the resources. The downward arrow linking behaviour with resources represents the important role that behaviour has in coordinating and reshaping resources. If an activity cannot be performed (e.g. contacting an expert or writing the report) because resources will not support it (e.g. insufficient contacts, insufficient domain knowledge or plans) then information-seeking might be performed to change this situation. Journalists manage and develop their internal and external resources to better define and support their task constraints. They learn, and in the context of memory limitations, manage external artefacts to reduce later referencing costs.

Influences of resources on constraints

The upward arrow linking resources with constraints represents the idea that a change in knowledge can lead to a reassessment or clarification of the constraints. Exclusivity checking, seeking evidence for an angle, and background information-seeking, for example, can develop the goals but can also provide insights to the journalist that suggest a new angle or show that the

angle must be changed. And changing how the constraint determinants are understood or defined, to a greater or lesser extent, changes the task. This influence, in combination with the influence that constraints have on behaviour and the influence that behaviour has on resources completes the cycle that was represented in figure 4.2.

Further, if information had been gathered according to the requirements of a superseded task definition, then that information can reduce in relevance and new information-seeking may be required. The combination of behaviour influencing constraints and then the reframed constraints influencing subsequent behaviour was observed in the exploratory study. Here, writing the report involved making commitments which changed the determinant of the *prior written commitments* constraint (the text as it is at any point in an assignment). This then leads to the initiation of new information-seeking. Hence unanticipated information needs arose once writing had begun and, in part, this created the need to relocate previously read source documents.

4.5.2 Corroborating and extending Bates, Ellis and Nicholas

4.5.2.1 Berrypicking-plus

Gathering information throughout searching, has been observed in both the exploratory and field studies, is central to Bates' (1989) Berrypicking model. In common with many information-seeking studies, the findings from the exploratory and field studies corroborate that model. However, what they add to this view of the information-seeking trajectory is the observation that users also often revisit documents to collect different information later. This aspect of information-seeking in the context of a wider task may have been outside Bates' original scope of concern. Perhaps, in the kind of scenario she was considering, documents would be printed and taken away from the information system and so supporting revisiting a document a number of times might be not be considered as a system design issue. But in the context of designing integrated systems that support both information-seeking and authoring, and on which some documents may only be read online with gathering performed using drag-and-drop functionality, then it clearly is a relevant issue for design.

4.5.2.2 A comparison with Ellis (1989a) and Nicholas & Martin (1997)

Comparisons can be made between the taxonomy of information behaviours arising from the field study, the information behaviour characteristics of social scientists found by Ellis (1989a), and the purposes to which information is put by national newspaper journalists as reported by Nicholas and Martin (1997). A comparison summary is shown in table 4.2 in which items judged as approximately equivalent by the author are shown aligned horizontally.

The table, showing the taxonomy from the field study in the centre column, Ellis (1989a) to the left and Nicholas and Martin (1997) to the right, is divided into three sections according to the three highest level categories of the field study model: information-seeking, information-gathering, and information reviewing. Two of the models, the one reported in this chapter and Ellis', are hierarchically structured, and, for these, lower-level items are indicated using an arrow.

(Ellis, 1989a)	**Information-seeking**	*(Nicholas & Martin, 1997)*
	Exclusivity checking	
Starting	Background info. seeking	To obtain context
→ Starter references	→ Seeking background overviews	
	Seeking evidence for a hypothesis	Researching
	Info. seeking for feature comparison	fact-checking (part of)
	→ Seeking properties of past disasters	
	→ Disc./conf. what someone said	
	Confirming names and their spellings	
	Identifying useful contacts	
Monitoring		Current awareness
		Stimulus
Chaining		
Browsing		
Differentiating		
Extracting		
Information-gathering		
	Dragging and dropping	
	Printing	
Information reviewing		
	Reviewing info. gathered during an assignment	
	Reviewing info. read but not gathered during an assignment	
	Reviewing information read prior to an assignment	

Table 4.2 A summary of comparisons between information behaviours reported in the field study, the information behaviour characteristics reported by Ellis (1989), and the 'purposes' reported by Nicholas and Martin (1997)

The table shows that the models partially overlap and can be though of as both corroborating and extending on each other. Probably the most striking difference between the models concerns their scope. Whereas Ellis, and Nicholas and Martin, were concerned only with information-seeking, the field study model, being more generally concerned with information behaviour, adds the categories *information-gathering* and *information reviewing*.

The model from the field study, and Nicholas and Martin's model are naturally very close since they are both based on data from the journalism task domain. Expressed from the perspective of the field study model, Nicholas and Martin report equivalents for *background information-seeking* and *seeking evidence for a hypothesis*. Also, Nicholas and Martin's general category of *fact-checking* can be thought of as broadly subsuming the more specific *information-seeking for feature comparison* (although the latter is not always as focussed as fact-checking would suggest). The model from the field study also includes the additional categories: *exclusivity checking*, *seeking background overviews*, *confirming names and their spellings* and *identifying useful contacts*. Similarly, Nicholas and Martin include the additional categories *current awareness* and *stimulus*.

In the field study model, the absence of a category matching *current awareness* (which matches Ellis' *monitoring*) can be explained with reference to the focus of the field study which is specifically concerned with behaviours that occur during news reporting and feature writing; monitoring for current awareness, in contrast, tends to occur during 'downtime'. Also, the absence of a category mapping onto Nicholas and Martin's *stimulus* can be best explained by the fact that, strictly speaking, this is an event rather than a behaviour. Although arguably, the idea of journalists feeding off unexpected finds, which is how Nicholas and Martin explain stimulus, is accounted for within in the field study model in relation to potential instability the *angle* and *accuracy* constraint determinants.

In comparing the field study model with Ellis' model, correspondence can also be argued between *starter references* (a sub-type or *starting*) and *seeking background overviews*. In both cases users seek out documents within their own domains which tend to provide good overviews. Categories within the field study model which do not appear in Ellis' model include *seeking evidence for a*

hypothesis, *information-seeking for feature comparison*, *seeking properties of past disasters*, *discovering/confirming what someone said*, *confirming names and their spellings* and *identifying useful contacts*. For the most part, these differences perhaps reflect differences between the tasks of journalists and social scientists. Categories included in Ellis' model which do not appear in the field study model are: *chaining*, *browsing*, *differentiating* and *extracting*. The absence of chaining (the following up of citations between documents) in the field study model can be explained by the absence of citations in the documents held within ENC service archives. Whilst *browsing*, *differentiating* and *extracting* undoubtedly are performed by journalists, they were not observed within the field study data.

4.5.3 The situated writer revisited

Towards the end of chapter 3, the results from the exploratory study were considered in terms of some of Suchman's Situated Action (Suchman, 1987), and in particular, her vague plan argument, which argues that few of our actions are explicitly planned, and where they are, plans are inherently vague. That, at best, plans provide only high level orientation, after which we respond to the contingencies of what real-life situations throw at us. She also makes the point that often it is only through engaging with a situation that its possibilities become clear, and that we do not know in detail the outcome of our activities when they begin.

These ideas were used to explain a number of findings from the exploratory study. In that study, despite an initial phase of intense information-seeking and gathering, during which the subjects were making judgements about what to write, new information needs arose once they engaged in writing itself. Some of these resulted in the desire to go back and gather different information from documents already read. This was described in chapter 3 as consistent with *some* pre-planning combined with a more reactive control mode in which the subjects ultimately both guided and responded to their own evolving texts.

In the current study, the idea of writing as uncertain, situated action, and of its plans evolving and changing in the face of the contingencies of a dynamic situation are, if anything, more prominent. The finding from the exploratory study of subjects reviewing previously seen documents to find information that they

hadn't gathered initially was also observed in the field study and this has been interpreted in terms of the more general cycle of behaviours changing constraints and then evolved constraints changing subsequent behaviour. The source of uncertainty begins with the initial indeterminate nature of and potential change in the task constraints, which was specifically emphasised in the analysis of constraints. The constraints were each explored for uncertain or dynamic elements through the identification of their *determinants*—factors which bind them to a particular meaning and a particular response in a particular assignment situation, and which can be unknown or change mid-task, and *determinant sources*—the origin of *determinant* binding. For this discussion, the determinants and sources for each constraint are summerised in table 4.3.

Constraint	Determinant	Source
Angle	angle	briefing
Deadline	deadline	briefing
Word-count	word-count	editor
Prior written commitments	the text	the writer's previous actions
Timeliness/currency	time since event	*many*
Proximity	readership and story	reader model, many
Exclusivity	angle & angles of previous stories	editor & cuttings archive
Human interest	human interest component	*many*
Historical context	historical context of story	many/cuttings archive
Accuracy	the facts of the story/spellings	*many*
Legal constraints	the law	training, colleagues *etc.*
Explanation	the events of the story	*many*
Cut from bottom	the importance of pieces of info	journalist
Original wording	the wording of articles	cuttings archive

Table 4.3 A summary of the determinants and sources for each constraint in the analysis

In light of the analysis of constraints, partial though the list was, it can be seen how the lab-study was a simplification, despite attempting to replicate the journalists' task situation. For example, in the exploratory study some of the constraints could be assumed or ignored, *viz.* timeliness/currency, proximity, exclusivity, legal; and the subjects' understanding of determinants of others remained static, *viz.* the angle, the deadline, the word-count and the facts of the story. Of the constraints that did apply in the exploratory study, the *prior written commitments* constraint has been used to explain the occurrence of information needs during writing, the re-initiation of information-seeking following the initial phase, and the need to relocate source documents. Since the determinant of this constraint is the text as it is at any point in an assignment, and since this evolves in often unplanned ways (*i.e.* situated action), new information needs occurred.

The findings from the field study, however, reveal greater uncertainty and potential for change. The angle, a key constraint of an assignment, can change, as can the required word-count and deadline. Further, encountering new information about a story can affect decisions about how to optimise the angle, proximity, human interest, accuracy, explanation, and cut-from-bottom constraints. In many cases this can affect the viability of an angle and suggest better alternatives. Likewise, given their sources, information-seeking from news cuttings archives can affect decisions about how to optimise the exclusivity, historical context, accuracy, explanation, cut-from-bottom and original wording constraints, and again, in many cases can affect the viability of an angle. Given the task constraints, encountered information is to the journalist what the currents are to Suchman's canoeist—the contingencies of the situation. Information-seeking is situated, reactive, and unpredictable; it is worth noting that within the Information Science literature, recognition of this is articulated as part of Bates' (1989) Berrypicking theory in which she argued that, during information-seeking, encounters with information can provide the user with new ideas and directions to follow

Determinant change and the gradual discovery of determinants through their various sources are at the heart of research and writing as a situated, uncertain and evolving process. For information behaviour, the implications of this are information-seeking in the first instance, the potential failure to make definitive relevance judgment when information is encountered, and the possibility of new information-seeking or the need to review documents read previously during an assignment in order to find more or different information.

This last behaviour will form a particular focus for this thesis from chapter 6 onwards. In chapter 6, a series of system requirements for integrated information retrieval and authoring systems are derived on the basis of the exploratory and field studies, and a requirement motivated by the need to easily review previously read documents is used to motivate the design of a prototype. The prototype combines novel hyper-linking functionality in connection with standard text copying by drag-and-drop. In chapter 7, this functionality is experimentally evaluated in comparison with traditional information-gathering (and review) techniques.

The next chapter broadens the task scope beyond journalistic news and feature writing to a general view of writing as a context for information behaviour. Based on the idea of writing as a kind of design activity it explores parallels between ideas emerging from research into the psychology of design and findings from research into information-seeking in the context of complex task performance (including the field study reported here). The chapter then uses the constraint perspective as a foundation for a framework for representing complex problems in a way that seeks to explain embedded information behaviour.

Chapter 5

Conceptualising the wider problem: a design psychology interpretation of writing tasks

5.1 Introduction

Having investigated journalist's information behaviours during writing assignments, it is now possible to address research question 3 by developing requirements for systems tailored to 'reflect' (Ellis, 1989a) the behaviours characterised by the model. This task, however, will be deferred until chapters 7 and 8 (following the current chapter) in which requirements will be explored, a prototype described, and its empirical evaluation reported.

Towards the end of chapter 4, the findings from the field study were summerised in the form of a model representing information behaviour within a rich description of its task context. This model sought to take a holistic perspective which would assign meaning to behaviours by locating them within a broader set of intentions, concerns and opportunities. Also, this model was essentially a representation of *process*—a process in which behaviours were understood as responding to and influencing an ecology of dynamic and evolving constraints and resources.

In the current chapter, which forms the basis for Attfield, Blandford and Dowell (2003b), some further theoretical perspectives are presented which contribute further to research question 2. These develop on the work reported in chapter 4 in two ways. First, the focus shifts somewhat from a consideration of process to a consideration of the task or 'problem situation' which lies at the heart of that process. The aim is to seek further understanding of information behaviour by considering more directly the problem from which it gets its motivation. Second, the focus generalises away from the specifics of journalistic writing to consider writing tasks in more general terms. The question being asked is 'what is the nature of the information seeker/writer's wider task?'

The latter shift relates to the idea of the grounded, inductive approach that was set out in chapter 1 of this thesis. In chapter 4, Grounded Theory method was used to construct a model by comparing and abstracting across some specific instances (represented within a set of interview accounts). By observing what was common across a number of cases, and combining cases into higher-order categories, concrete instances were reduced more essential, generalised ideas.

The value of such an approach is that, at all times, abstraction remains firmly grounded within, and hence faithful to, a set of acquired raw data. In chapter 1, it was argued that, as abstraction from particular observations, theory permits findings to be evaluated in terms of new situations. Hence, research communities can collaborate around and develop common sets of ideas, providing coherence to what might otherwise be disparate research efforts. This corresponds to an approach for studying purposeful action that Suchman identified as having emerged from Anthropology and Sociology, which recommends "... building generalisations inductively from records of particular, naturally occurring activities, and maintaining the theory's accountability to that evidence." (Suchman, 1987, p179).

By comparing (and contrasting) findings across different study situations new insights can emerge. There is no reason, however, why the process of comparison (and abstraction) should remain within a single discipline. In a sense, this chapter follows an established approach to information-seeking research in which concepts and ideas from related areas have been applied in order to enhance existing models and provide theoretical leverage. Wilson (1996) for example, makes a strong case for this kind of cross-disciplinary theorising. Observing that Information Science was only one of a number of fields containing work relevant to the study of information-seeking behaviour, Wilson considered research derived from a number of areas *viz.* the study of personality in psychology; consumer behaviour; innovation research; health communication studies; organizational decision-making; and information requirements in information systems design, and related these to his own model of information behaviour.

In the current chapter, elements of the field study findings are disassociated from their specific context through relationships drawn with other research findings in Information Science and also findings in Design Psychology. Central to this is the view that writing in general can usefully be conceptualised as a form of design activity and that, by taking this perspective, we are better able to interpret the information-seeking behaviour of authors.' The link between writing and design has been made elsewhere. For example, Goel and Pirolli (1992), in their analysis of design problem structures, identify writing as diverging only slightly from prototypical design tasks such as architecture and engineering. Moreover, central to Sharples (1996) model of writing is the idea of the writer as

a creative designer. But the significance of this idea for information-seeking research is that it can provide additional leverage in understanding the information-seeking phenomena that arise in the context of writing tasks.

The aims of this chapter are twofold: the first is to identify and explore parallels between the findings in the separate areas of the psychology of design and information-seeking in the context of complex task performance. It is shown that significant parallels can be identified with a related, but for the Information Scientist, perhaps unfamiliar, branch of research. The value in drawing these parallels is that explanations offered within the psychology of design literature can be applied to explain a number of information-seeking phenomena. The second aim is to develop a design-based representation of writing tasks as a means of providing a situated account of phenomena such as information-seeking uncertainty, the progressive refinement of information-seeking focus, and the reciprocal relationship between a user's evolving conception of their task and the information that they find. Specifically, the idea of a constraint delimited problem space is introduced as the basis for a framework for representing the information seeker/writer's 'problematic situation' (Wersig, 1979).

To provide context, the chapter begins by summarising some of the contributions to information-seeking research reviewed in chapter 2 which will be useful for making the case for parallels with design psychology and for showing the value of the framework. Important themes in this work are uncertainty, its relationship with the formulation of a task focus, and the effect that this has on relevance judgements and query specificity.

These summaries prepare the ground for the fourth section in which four features of design problems and design problem solving as observed within the design psychology literature are discussed, and related to concepts in information-seeking. These features are: incomplete specification; primary generators; the analysis/synthesis dynamic; and multiple constraints and integrated solutions. These concepts are exemplified using observations from information-seeking research in general also with reference to findings from the field study reported in chapter 5. Hence, this chapter provides a theoretical development of that study. Finally, the design perspective is used to motivate a constraint-based framework for the representation of writing tasks which it is

claimed offers a new explanatory framework for interpreting many information-seeking phenomena.

5.2 Revisiting uncertainty, formulation and problem solving

Uncertainty on the part of users engaged in information-seeking has gained increasing prominence as an issue for user-centred information-seeking research. Failing to accommodate user uncertainty is often cited as a shortcoming of the systems oriented (Cranfield) approach in IR research which, it has been argued, is based on assumptions of certainty and order (Belkin, Oddy & Brooks, 1982; Kuhlthau, 1999).

According to Belkin *et al.* (1982a), there are times when a user is able to specify what information they require, but more usually the information that is required cannot be clearly specified in advance. Influenced by constructivism, and based on a series of studies of students and other novice library users, Kuhlthau's ISP model identifies six stages of the information search process through which an information seeker moves on the path from uncertainty to a constructed understanding. The most critical part of the ISP process is the point where the information seeker forms a focus for their task (formulation); this acts as a turning point. Following formulation thoughts become clearer, uncertainty gives way to confidence, and clarity and confidence increase as the user gathers information (collection). Finally, a sense of relief is experienced as the search is completed (presentation).

In a recent study, Kuhlthau and Tama (2001) investigated the applicability of the model to the information-seeking processes of lawyers undertaking a range of tasks, including both complex and routine tasks. This study set out to assess, among other things, whether, in this task domain, higher levels of uncertainty and construction of new knowledge were associated with more complex tasks. The findings supported this relation and accorded closely with the ISP model.

In the case of lawyers, complex tasks (preparing a case for trial) required considerable thinking and formulation. Formulation corresponded with establishing a trial strategy and was described as a difficult but creative part of

the process. Further, developing a strategy, as well as being associated with uncertainty about the wider task, was also associated with uncertainty concerning the information needed. As predicted by the ISP model, the lawyers initially sought overview and background information to assist formulating a strategy.

Chapter 2 outlined the acceptance of the idea of formulation as a particularly significant point for information-seeking. Vakkari interpreted it as a pivotal point when a perspective or hypothesis is created on the wider task (Vakkari, 2001). For Byström and Järvelin (1995) it is the creation of a solution space which determines the information requirements of the task. After formulation the user has a problem that might be solved, and knows more clearly what information they want (Vakkari, 1999).

In chapter 2, a number of studies were also reviewed which explored the progressive reduction of uncertainty as indicated by increasing query specificity and increasingly categorical relevance judgements. Tang and Solomon (1998) showed how a student's relevance criteria became more focused during the review of a bibliographic results set for writing a term paper. Further, after she had read her selected papers, she changed the topic to one which more closely corresponded to the content of the retrieved documents.

In Yang's (1997) study of undergraduates seeking information in order to write a class assignment he noted that subjects were sometimes uncertain about the value of information they encountered and would defer judgement to a later time. Yang also observed one subject, at an early stage of his task, exploring available information in the hope of hitting on something that might trigger an insight or idea (Yang, 1997).

In a four month longitudinal study of the information-seeking of students engaged in the task of writing a research proposal for their masters theses, Vakkari *et al.* (reported in Vakkari, 2000a; Vakkari & Hakala, 2000; Vakkari, 2000b; Vakkari & Pennanen, 2000; summarised in Vakkari, 2001), condensed Kuhlthau's six stages into three: prefocus, focus and postfocus and observed a systematic relationship between task stages and the evolving polarity of relevance judgements, and increasing query specificity.

Vakkari's work, like others, places an emphasis on the significance of developing a focus for a wider task in relation to information-seeking behaviour. He describes formulating a focus or a guiding idea as a critical, pivotal point when a particular perspective is formed and the user moves out of uncertainty to understanding (Vakkari, 2001). Using *a priori* determinability as a signature characteristic of task complexity, Byström and Järvelin classified tasks according to uncertainty concerning task outcomes, process and information requirements and described focus formulation as creating a solution space and in doing so reducing the uncertainty associated with a task's information requirements.

Two cognitive issues can be understood as being involved in formulation: first, the information seeker achieves a pivotal level of sufficiency in their understanding about a topic; second, that understanding is sufficient to support the formulation of a focus for a wider task. Like Vakkari's notion of formulation being comparable to the generation of a hypothesis for accomplishing the wider task, a focus can be interpreted as corresponding to the development of a broad plan and this reduces uncertainty about the information needed. No doubt the plan lacks detail, but it is nevertheless more focused than any original statement of objectives—the goal that brought them to the information service in the first place.

By developing a focus, or guiding idea, a searcher creates a solution space *i.e.* clearer task goals (a reduction in task uncertainty), and so their information requirements become clearer (a reduction in information-seeking uncertainty), and this manifests itself through the production of more specific queries and the ability to make more confident and discriminating relevance judgements.

5.3 The author as a designer of text

In this section the problem solving view is extended by presenting writing not simply as a kind of problem solving, but specifically, as design problem solving. Drawing on influential work in the psychology of design (Lawson, 1997; Darke, 1978; Goel and Pirolli, 1992; Schön, 1983), some characteristic features of design problems and design problem solving are reviewed which are related to features of writing, extending earlier work by Sharples (1996), and also to features of information-seeking as reported in the Information Science literature.

Many phenomena are illustrated using descriptions of the information behaviours of newspaper journalists reported in chapter 5. The design problem features focussed on are: incomplete specification; primary generators; the analysis/synthesis dynamic; and multiple constraints and integrated solutions.

5.3.1 Incomplete specification

Incomplete specification as a feature of design problems was recognised initially by Reitman (1964) (under the label of ill-defined or ill-structured problems) and has subsequently been echoed throughout the psychology of design literature (see, for example, Goel & Pirolli, 1992; Lawson, 1997). Reitman noted that there exists a lack of information (*i.e.* there is ambiguity) in the three components that comprise design problems: the start state, the goal state, and the transformation function from the start to goal states. Similarly, Sharples (1996) relates this feature of design problems to writing tasks. Unlike the classic problems studied by cognitive psychologists, like chess or Towers of Hanoi, there is no fixed set of goals or sequence of steps for solving them (Sharples, 1996).

A clear relationship can be seen between the notion of incomplete specification in design and that of *a priori* indeterminability concerning information requirements, process and task outcomes used by Byström and Järvelin and others. Also, Byström and Järvelin argue that *a priori* determinability is relative to the point of view of the user, and the same view is expressed by design psychology researchers with respect to indeterminability. For example, Goel and Pirolli (1992) cite Simon (1973), who argued that a problem is not intrinsically unstructured, but that this is a function of the relationship between the problem solver, their available knowledge, and the problem to be solved. Lawson demonstrates this point using the example of igloo building, arguing that this is not a design problem at all for an Eskimo (*i.e.* an expert), but rather a traditional form of solution or 'vernacular' (Lawson, 1997) with variations to suit different circumstances. In the context of writing research, Sharples (1996) explains this relativity with the idea that expert writers can call on a large stock of remembered plans and schemas built up through a long apprenticeship in the craft of writing, whereas inexpert writers have less pre-compiled knowledge and so must construct plans to order.

News journalism, being a particularly formulaic genre of writing, is replete with such schemas determining either structure or content, such as the schemas specified by the cut-from-bottom constraint or the explanation constraint. As one specialist correspondent explained:

> **MG:** Well people, yeah, well people want explanations. They want to know why—why is there an outbreak of foot and mouth? why are we.. you know why are we getting CJD?

Comparing casualty figures and the level of emergency response were also described as standard practice for disaster reporting as metrics of severity.

Clearly, for the experienced journalist, schemas supplement the initial brief and reduce task indeterminacy. And since *a priori* indeterminacy corresponds with information need uncertainty, we can expect the experienced journalist to know better what information to search for than the novice.

Within the design literature this relationship between incomplete specification and information need uncertainty has been commented on by Lawson, who observes that, given incomplete specification, it is difficult for designers to know what problems are relevant and what information will be useful until a solution is attempted (Lawson, 1997). Moreover, Lawson regards the ability to live with this uncertainty as an important personal quality for a designer, and he criticises modern Computer Aided Design systems for failing to accommodate uncertainty, particularly during the early stages of the design process. This claim has a notable resonance with the ideas of Kuhlthau who has long argued that bibliographic information retrieval systems are ill-suited to users in a state of uncertainty (Kuhlthau, 1993).

Incomplete specification means that the problem itself is not apparent but must be found (Lawson, 1997). Consequently, much of a designer's time is spent in identifying and refining the problem (Sharples, 1996). Goel and Pirolli (1992) argue that lack of specification in design problems means that extensive problem structuring must be performed. Byström and Järvelin (1995) echo this in the context of complex tasks with embedded information-seeking.

In the next section the relationship between problem structuring in design tasks through 'primary generators' (as described within the psychology of design) and problem structuring through focus formulation (as described within the information-seeking literature) is discussed.

5.3.2 Primary generators

The notion of a *primary generator* as a means of structuring design problems has become particularly important within the psychology of design literature. A *primary generator* is described as a simple but powerful idea or principle established by a designer early in the design process around which further design activities are subsequently organised. The idea is attributed to Darke (1978) who interviewed a series of architects about their intentions when designing local authority housing. Darke found that the architects latched on to a relatively simple idea early on, and that this idea would then narrow down the space of possible solutions by providing an initial focus *i.e.* by constraining and guiding the designer's development of a solution. Darke gives examples of primary generators, including the idea of creating a mews type street, or the idea of leaving as much open space as possible. Lawson (1997) observed that some designers deliberately generate a series of alternative primary generators, followed by progressive refinement, testing and selection.

Sharples (1996) incorporated the idea of primary generators into his model of writing, noting that accomplished writers often describe specific concepts and ideas as initiating their writing. For example, he cites Garcia Marquez who explained that the writing of One Hundred Years of Solitude was organized around the adoption of a particular tone.

In the field study reported in chapter 5, it was reported that all news assignments adopt an explicitly articulated angle, which is usually communicated to the journalist by their editor during an initial assignment brief. It was also reported that an angle, perhaps contrary to popular perception, is not so much an emotive stance or value judgement, but rather takes the form of a proposition, or central factual claim that is to be made by the report. And where the claim involves some speculation, the angle takes the form of a working *hypothesis* or *conjecture.* It is then tested by information-seeking to either confirm or refute it. As the Chief Sub-Editor quoted in section 5.5 said:

> **QC:** ... Essentially there is an *angle* to all news and features; it is really a working hypothesis that translates the gathered *facts*, which may include some speculation, into a coherent account.

An angle, then, appears to correspond with Darke's notion of a primary generator. It is an idea that is developed early in an assignment around which further activities are subsequently organised. By providing the journalist with a focus, the angle narrows down the space of possible solutions. Also, as an early, focused perspective or guiding idea which determines both a solution space and the writer's information requirements, it similarly illustrates an assignment *focus* as discussed in the information-seeking literature. Establishing a primary generator represents a form of focus formulation applicable to writing with and without embedded information-seeking.

Observations in the psychology of design relating to the potential instability of the primary generator allow us to explore additional similarities. Both Lawson (1997) and Sharples (1996) point out that as a design process progresses, designers can gain new insights into their problem, leading them to reject or modify an initial primary generator (the idea through which the insight was achieved). Similarly, in his study of undergraduates' relevance judgements Yang reported that, "Although each subject established a framework to guide his or her problem-solving, these were treated as malleable and open to change" (Yang, 1997, p.86) and "a search might trigger a competing or more compelling idea, enabling or enticing the subjects to change direction, set aside, forsake, or even forget the original focus of their search" (Yang, 1997, p.81). And in their case study of the graduate student searching for information in preparation for a term paper, Tang and Solomon observed a change in focus and concluded that, "The shift to a new topic came about through the subject's interaction with the documents and her learning about the issues that were of concern to authors of texts that were related to her interests within the course situation" (Tang & Solomon, 1998, p.254). This same idea was central to Bates' (1989) Berrypicking model and the idea of evolving search. Finally, in chapter 5, it was reported that journalists describe the process of information-seeking and writing as sometimes enabling a better idea for an angle than the initial idea, and hence destabilising the process. CM had said:

> **CM** Writing plans can change 'relatively frequently'. You might come up with what you think is a better idea than the one that the news desk gave you. You've then got to, of course, then got to go and convince the newsdesk that you've got a better idea than the one they gave you. Which is not always easy. But you know, its something that should be encouraged. If you stumble across something that's more interesting than the original line that you've got then you've got to change it...

Modifying an initial idea for a writing task, as a result of the information-seeking that was motivated by that initial idea was represented in Chapter 5 (Figure 5.3) as a cycle. It arises from the situated fluidity of writing with embedded information-seeking and is one way in which such activities can usefully be considered as part of a design process.

In the next section further parallels between design problem solving and writing with embedded information-seeking are identified through the description of a prominent conceptualisation of the cognitive processes of design activities—the analysis/synthesis dynamic. This description will contribute to an explanation of the instability discussed in the latter part of this section.

5.3.3 The analysis/synthesis dynamic

Given the fluid unpredictability of design, there has been a trend within the psychology of design literature away from models representing the design process at the macroscopic level as a linear sequence of activity phases, in favour of models which represent iterations between the fundamental activities of analysis and synthesis. Schön, who suggested that this dynamic is at the heart of all design professions, referred to it as "a reflective conversation with the situation" (Schön, 1983, p.76). The designer engages in a continual process of making a move, reviewing the situation to assess the result of the move, and moving again; and during a review they may form a new appreciation of their problem. Schön refers to this as the situation 'talking-back' to the designer and the designer responding and frequently re-construing the problem; a process consisting of an interplay between exploration and commitment rather than a series of predetermined moves. Further, Schön identifies this process as occurring at different levels of granularity as the designer shifts focus backward and forward between the unit and the whole.

Schön's idea of a conversation with the situation is precisely what Suchman meant by her notion of *situated action* (reviewed in chapter 2). Suchman argued that,

> It is frequently only on acting in a present situation that its possibilities become clear, and we often do not know ahead of time, or at least not with any specificity, what future state we desire to bring about
>
> **(Suchman, 1994, p.52)**

Generally speaking, for Suchman, the specificity of future states only become realised when vague intentions are played out against the contingencies of an unpredictable environment (Suchman, 1994). In the case of designing, the unpredictable contingencies are the way things turn out or the things that are discovered following a move. In the case of information-seeking and writing, a move could be the posing of a query or the writing of a sentence—the effects of these being potentially as unpredictable as the water currents of the rapids are to Suchman's canoeist. Analysis of the new situation and its relation to whatever count as desirable outcomes can then be reviewed and subsequent moves chosen.

Lawson refers to the analysis/synthesis dynamic as "analysis through synthesis" (Lawson, 1997, p.43), and places particular emphasis on the idea of learning through experimentation. He describes a study to explore cognitive styles in solving design problems. Two groups of students, science students and architecture students, were given a problem that required them to create a structure from wooden blocks according to a set of specified constraints. An apparent difference between the two groups was that the science students attempted an *a priori* analysis of the problem in search of a rule for constructing an optimal solution. The architecture students, on the other hand, consistently used the strategy of learning about the problem through attempts to create solutions. The analysis through synthesis dynamic was also reported by Eastman (1970) as a result of observations of experienced designers redesigning a bathroom, and by Akin (1986), who observed a group of architects designing buildings. Akin found that the architects constantly generated new goals and redefined constraints, a process whereby the designer discovers more about the problem as they critically evaluate their own solutions (Lawson, 1997).

An analysis/synthesis dynamic is evident within Sharples' (1996) model of writing which emphasises the interplay between text production (engagement), and reflection on the text produced (reflection). During engagement, the writer is devoted to the task of turning ideas into text. During reflection, the writer reads the text (reviewing), forms new ideas (contemplation), and makes decisions about what else to write and how to organise it (planning) (see figure 2.6). In chapter 2, information-seeking was identified as an analytic activity that, in Sharples model, would most ideally be located within the contemplation sub-process. As with the distinction between the science and architecture students observed by Lawson, however, different people may have different writing styles relating to the extent to which they pre-plan writing or make decisions as they go. The advantage of an analysis/synthesis model, as Sharples points out, is that it can account for such differences through variations in the attention users devote to each process throughout the cycle

Writing is a dynamic and emergent activity. Notwithstanding pre-compiled schemas, many decisions about content and structure are made as the writer progresses towards a solution and understands better what it is they are writing. This was shown through the provisional and deferred relevance judgements in the Yang (1997) study, and changes in relevance judgements in Tang and Solomon's (1998) study. Similarly, the point is illustrated by the observation that journalists often modify their initial idea in response to found information, and, that as the journalist works on an assignment their concept of the detail to include is constantly being shaped and reshaped. As CJ explained:

> **CJ:** ...the ideas will take shape all the time... at the point that they change all the time. It is only really when you have to sit down and actually write it that I would have to decide what way to go into the story... I am preparing it... I am preparing all the time.

In other words much of the activity is situated (in Suchman's sense). It is ongoing and reactive to the opportunities presented by the situation.

The significance of the analysis through synthesis dynamic for information-seeking in the context of writing is that, through this dynamic, the writer engages in a continual process of making, reviewing and adjusting commitments. As

each commitment is made, so this constrains and determines subsequent commitments which ultimately contribute to a coherent whole. Effectively, the problem emerges with the solution. As Dewey argued, "we know what the problem exactly is simultaneously with finding a way out and getting it resolved. Problem and solution stand out completely at the same time. Up to that point, our grasp of the problem has been more or less vague and tentative" (cited from Kuhlthau, 1993, p.341). And where a commitment, or decision, has implications for particular information requirements, so a new information need arises.

Some models of information-seeking and information behaviour, including Kuhlthau's ISP model, and Wilson's (1999) problem solving model, have adopted a representation depicting a linear sequence of activities in which the user progresses from establishing a problem, to refining a solution. Others have avoided commitments to linearity (see for example Ellis & Haugan, 1997). Linear models, it can be argued, run the risk of under-representing the indeterminate nature of complex problems and the consequent twists and turns of human exploration and creativity; this risk is perhaps what underlies the feedback loops shown in Wilson's model.

So far three features of design problems and design problem solving have been considered: incomplete specification, primary generators, and the analysis/synthesis dynamic, and these have been related in particular to information-seeking uncertainty and focus formulation. A primary generator acts as a constraint to reduce the problem space, but there are many other constraints that collectively structure a design problem. A primary generator, albeit a constraint that is psychologically significant to the design process, is nevertheless one of many constraints. In the next section the review of design problem features is completed with a discussion of the idea that design problems require integrated solutions to multiple constraints. Indeed, as Lawson argued, a design problem *is* the sum total of its constraints. This will have important implications for the representation of the information seeker/writers problematic situation presented later.

5.3.4 Multiple Constraints and Integrated Solutions

According to Lawson (1997), a constraint is an issue that must be taken into account when forming a solution to a problem. Good design is frequently an

integrated solution to a whole cluster of constraints. Lawson describes design problems as being "built up" or constituted of constraints (Lawson, 1997, p100). In an exploration of the types of constraint that can be part of a design problem, he proposes a three dimensional model on which all constraints can be classified, which together provides a general framework for differentiating different kinds of design problems. For example, one of Lawson's dimensions, which is also discussed by Goel and Pirolli (1992), corresponds with the extent to which a constraint is *hard* or *soft.* Hard constraints are rigid and must be satisfied, but design problems are frequently constituted mostly from *soft* constraints; these are less rigid and render the problem more as one of constraint optimisation rather than constraint satisfaction.

Although this model is only referenced in passing here, such an endeavour is clearly underpinned by the premise that a design problem *is* the aggregate of its constraints, and that constraints provide the designer with a problem space within which one or more solutions lie. However, to complicate the matter, the requirement for an integrated solution often leads to constraints acting against each other, and trade-offs must be made.

Writing *is* such a multiple-constraint problem, as noted by Sharples (1996) and, as with any task, the constraints at play in writing characteristically vary between types of task and differentiate one task from another. Since constraints constitute the task, constraint differences differentiate one task from another. The constraints of writing are complex and operate at many levels, ranging from issues of surface level structure to social and political implications. Among the many constraints reported in chapter 5 as providing the context for the work of journalists were included: the angle (which typically serves as both a primary generator and a constraint), deadline, word count, proximity, exclusivity (that the story angle should not repeat an angle taken in any previous article), accuracy, legal constraints, and cut-from-bottom structure (that important information should appear before less important information). These multiple constraints are satisfied in the design of a good integrated solution—namely, the writing of a publishable newspaper article.

In elaborating constraint types in writing, Sharples makes a further distinction (adopted from Lawson (1997)) between constraints which are *external* to the writer, such as an essay topic, previously written material, or a set of publisher's

guidelines, and those which are *internal* to the writer, such as schemas, inter-related concepts, genres and knowledge of language. As with Lawson (1997), Sharples identifies the resources a writer uses as also constraining the writing process. "The designer, the artefacts, and the setting form a rich interoperative system. Each artefact conditions the activity, assisting certain operations while restricting others" (Sharples, 1996, p.11); it is in this context that he makes passing reference to information-seeking from external resources.

One important observation by Sharples (1996) is the apparent paradox that constraints are, simultaneously, limiting and facilitating. They are limiting in that they define the space of acceptable solutions, and yet they are facilitating insofar as they enable the creative process by "constraining the generative system into an appropriate conceptual space" (Sharples, 1996, p.3) *i.e.* they provide focus. The way that a writer generates new material, and also manages the proliferation of possible next actions, is by imposing appropriate constraints. Put simply, constraints, as well as limiting the design process, guide the designer.

5.4 A constraint-based account of writing: explaining embedded information behaviour

Where information-seeking is embedded within writing, a reciprocal relationship occurs between them. Information needs are determined by the needs of the task, and yet the evolving task is shaped by the information found; this is an aspect of the analysis/synthesis dynamic. Consequently, a representation of the wider task which reflects this view should explain how tasks give rise to information needs and also how found information affects task structure. Further, on the view that a design problem *is* the totality of its constraints, a representation of the wider task should use constraints as its conceptual basis. And, since design problems feature multiple constraints and require integrated solutions, it should also show that successful solutions do just this. In this section these requirements are used as the point of departure for a representation of writing tasks.

The representation of writing tasks developed is based on the idea of a task as the aggregate of its constraints using news-writing as an example. The idea that a news report or feature article must satisfy, or at least optimise, multiple constraints has been discussed. For the sake of brevity, however, the framework is developed in terms of only three: originality, truth and newsworthiness. In chapter 4, it was reported that these three constraints are, in particular, applicable to the determination of a good angle at the inception of an assignment (as well as throughout) and this is the context in which they will be used to exemplify the framework. This simple account may well underestimate the number of constraints used to determine a good angle and certainly underestimates the constraints acting on a finished report. Nevertheless, the representation and the conception it offers of a writing task can be extended to accommodate any number of constraints.

The framework is shown as a constraint delimited problem space in figure 5.1. This represents the idea that each constraint independently defines its own space of satisficing solutions. For example, the originality constraint defines news report ideas that are *new* but not necessarily *true* or *interesting* since they may fall outside the boundary of the accuracy and newsworthiness constraints. Since, for a good solution to the problem of determining a story angle all three constraints should be optimised, integrated solutions lie at the intersection of all three. In effect, any idea falling within this intersection represents a viable assignment opportunity. Using this framework, the initiation and effects of information-seeking through two scenarios will be explored using journalistic writing as an example. (For the sake of ease of representation and clarity, we have shown each of the constraints as having clearly defined boundaries, *i.e.* they are represented as *hard* constraints, ignoring the possibility that any of them may be *soft*.)

Consider a journalist at the very earliest stage of an assignment—before an angle (the 'primary generator' of a story) has been established. Indeed, at this point it can be said that an assignment does not yet exist. Let us say that she is a senior, specialist journalist who has the authority to source her own stories. Information-seeking at this stage might typically be broad-based with unspecified, or difficult to specify, needs, and might typically feature monitoring activities such as reading incoming newswires and emails, and receiving telephone calls. She may even use more proactive means such as contacting

specific agencies to see whether anything interesting is in the offing, but at the moment a story does not exist.

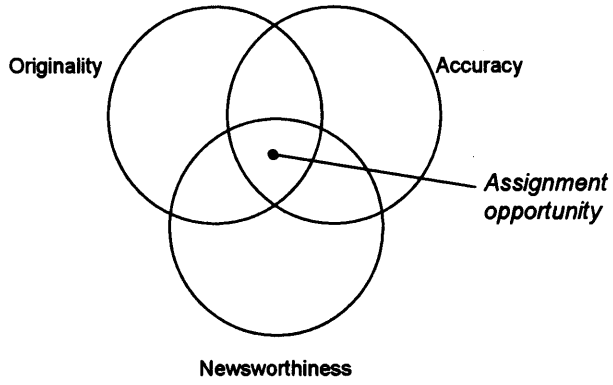


Figure 5.1. For a journalist, a good angle (story idea) lies at the intersection of the constraints of *originality*, *accuracy* and *newsworthiness*.

Although she is aware of the constraints represented in figure 5.1, at present, the journalist has no idea for a story that will meet them. Through monitoring, however, some information comes to her attention which triggers an idea (*i.e.* an angle, or primary generator) for a solution which might optimise the constraints. The concept for the story may not simply be reporting the information received—it might involve an inference drawn from that information—but, whichever the case, she judges that the resulting story angle is newsworthy. However, some doubt exists in her mind about how close to the truth the central claim, or angle, of the story would be. Uncertainty also exists about the originality of the idea. Consequently, she engages in information-seeking in order to resolve these (and other) issues.

Figure 5.2 represents these developments in terms of the problem space understood from the journalist's perspective. In the initial state (left), the journalist has a problem space but no competing solutions. When the new information arrives, it triggers a primary generator for a story and, as a proposed solution, this can be located within the problem space (shown right). This idea (marked 'S') represents an opportunity and so provides the journalist with focus.

However, although the journalist judges the proposal newsworthy, and hence it is shown within the boundary of the *newsworthiness* constraint, she is unsure about where it is located in relation to the boundaries of the *originality* and

accuracy constraints. She is unsure of the constraint boundaries. In Figure 5.2 this situation is represented (right) by showing the *originality* and *accuracy* constraints with two alternative perimeters (dashed lines). The proposed solution is located within the problem space, but its position in relation to the *originality* and *accuracy* constraints is indeterminate.

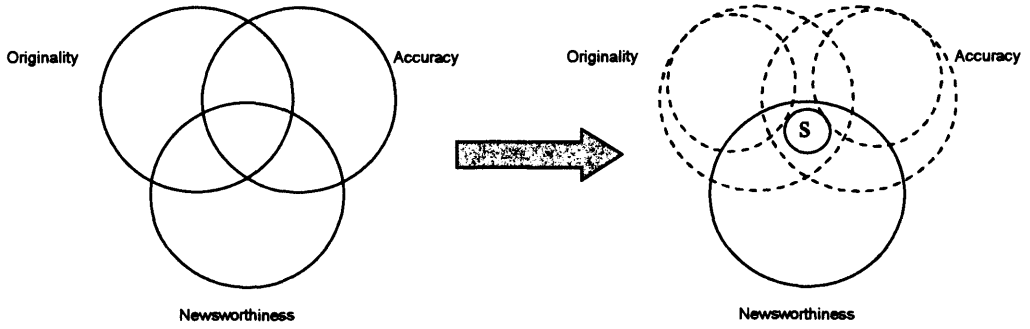


Figure 5.2. A solution space representation of the transition from no story concept to a story concept with originality and correspondence uncertainty.

The reason for showing the story idea in figure 5.2 as a circle is to indicate that, whilst this idea is in embryonic form (*i.e.* as an initial focus, angle, vague intention or primary generator), rather than representing a single solution, it represents a class of solutions, *i.e.* all those solutions that adopt the given angle. Hence, the idea itself constrains (or focuses) the problem space. By providing focus, it has the effect of narrowing the space of possible solutions by "constraining the generative system into an appropriate conceptual space" (Sharples, 1996, p.3). Adopting a primary generator, however, is only the first step in narrowing the problem space. With each subsequent commitment the journalist will further reduce the size of the available solution class, until ultimately it consists of a single solution.

The second scenario features the same journalist at a later stage in her assignment, when she has nearly finished writing. Her writing has been occasionally interrupted by information-seeking. During cycles of reflection (analysis) on what she has been writing (synthesis) she has identified new information needs and has interrupted writing to resolve them. Occasionally she writes something, reflects and then changes her mind. During one such cycle, she reflects on the fact that, given what she has written already, she ought to

provide the reader with a date as part of the background information. She believes she knows this information and has recorded it somewhere, but on reflection she is not entirely confident. We will refer to this piece of information as proposition *p*. The journalist directs a request to an information channel that she considers sufficiently reliable, and awaits a response.

This information need has arisen because the journalist is proposing a particular solution to her assignment, *viz.* one that incorporates proposition *p,* but she is unsure whether *p* is true. In terms of the constraint delimited problem space, she is fairly confident where the solution lies in relation to the *originality* and *newsworthiness* constraints, but she is unsure of its relationship with the boundary of the *accuracy* constraint. This uncertainty is represented in figure 5.3 as the *accuracy* constraint having two alternative perimeters (dashed lines) with one possibility incorporating the solution and one not. Also, since at this point in the assignment the proposed solution is far more refined than the solution class of the previous example, it is shown as a single point.

By confirming or disconfirming *p*, the journalist resolves this uncertainty. To confirm *p* is to establish that the contribution of *p* is not to render the solution outside of the constraints. To find that *p* is false is to establish that the solution falls beyond the space of acceptable solutions. In this case, *p* would need to be modified or dropped altogether, and where the journalist fails to find out either way, the uncertainty remains. A response to this situation can be to dilute the claim and, in doing so, raise the certainty that the solution falls within the intersection of all the constraints, although this might reduce the value of the report in terms of other constraints, such as newsworthiness.

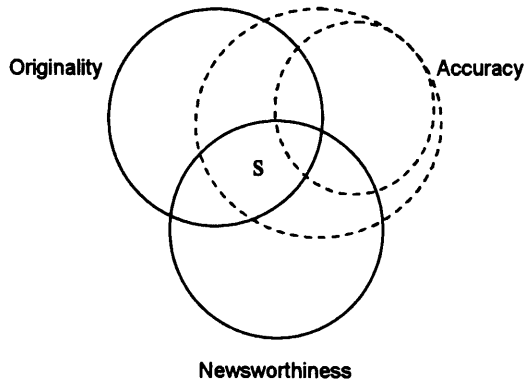


Figure 5.3. Accuracy uncertainty represented within a news assignment constraint space

These two examples illustrate not only how information-seeking can change a writer's understanding of the constraints that define their task, but also how framing a writing task in terms of a constraint delimited problem space can account for information-seeking in the first place. Hence, the model implies the situated cycle of a task giving rise to information-seeking and the information found changing the nature of the task *etc.*

The examples intentionally demonstrate two contrasting information need types: a broad need specified at a general level, and a well specified fact-checking need. In the first case, the broad need arises from a requirement for a focus or primary generator; the writer is in a state of uncertainty. This state is represented as a problem space with no candidate solutions. Information-seeking then provides the writer with an *opportunity* by triggering a solution idea or, rather, an idea of a class of solutions which, it is hoped, fall within the optimal area of the problem space. Subsequent information-seeking can then resolve uncertainty with respect to the relationship between the proposed solution class and other constraint boundaries. In the second scenario, a well-developed solution gives rise to a well-specified need—again, in order to resolve uncertainty with regard to the relative location of solution and the boundary of the accuracy constraint.

Hence, information-seeking can give shape to the problem space in (at least) two ways. On the one hand, it can reshape the problem space by enabling the information seeker to identify an opportunity, which, in turn, better defines their problem and so establishes new constraints. On the other hand, it can reveal the shape of existing constraints and, in particular, how their boundaries correspond with different solution proposals. Design problems are typically under-specified at the outset, becoming better specified as solutions are attempted. The idea of constraints emerging during and through task performance is characteristic of design problem solving and, in particular, of the analysis/synthesis dynamic. Since the constraints *are* the problem, commitments which change the constraints effectively change what the problem is. Efforts to explore the existing constraints in relation to a given solution proposal change the structure of the constraints as understood by the problem solver.

5.5 Summary and Discussion

The first aim of this chapter was to identify and explore parallels between what researchers studying the psychology of design have found and what information scientists have found in studies of information-seeking in the context of complex information tasks, with a particular focus on the task of writing. Four features of design problems and design problem solving were reviewed, and these were related to existing theories and empirical findings in information-seeking.

In section 5.3.1, it was argued that the idea of design problems being radically under-specified, and therefore requiring significant structuring, corresponds with Byström and Järvelin's notion of *genuine decision tasks.* Byström and Järvelin (1995) argued that *a priori* indeterminability is relative to the point of view of the user, and we find the same idea within the psychology of design in relation to user expertise. We reviewed an explanation of this from the psychology of design literature made in terms of learned plans and schemas.

In section 5.3.2, it was argued that structuring a problem by establishing a primary generator, as identified in the psychology of design literature, corresponds with the idea of finding a focus in complex information tasks as explored in the work of Kuhlthau, Vakkari, Byström and Järvelin and others; a primary generator being an imposed constraint that narrows the space of potential solutions and, in doing so, focuses and guides the user's concept of what information is and is not relevant to the task. The idea of instability of the primary generator was accordingly related to focus reformulation.

In section 5.3.3, it was argued that a conception of process which is prominent within the psychology of design literature, and which is referred to here as the analysis/synthesis dynamic, offers an intuitive explanation for the evolution of and changes in focus. Analysis and synthesis implies that users move and then evaluate their new situation *i.e.* they have a conversation with the situation. This was related to Suchman's *situated action* and the idea of people, equipped with vague plans, responding to the contingencies of an unpredictable situation. It was argued that the analysis/synthesis dynamic provides an account of the dynamic nature of human exploration and creativity more adequately than linear activity sequence models.

In addition to the primary generator, writing tasks are driven by a need to optimise multiple external and internal constraints, and in section 5.3.4 the idea of design problems being constituted from constraints was related to writing tasks with embedded information-seeking.

The second aim of this chapter was to develop a conceptualisation of the information seeker/writer's 'problematic situation' based around the ideas reviewed in section 5.3 and which would provide a framework for explaining various aspects of information-seeking behaviour. In section 5.4, a framework was developed for representing writing tasks in terms of a constraint delimited problem space, illustrated using findings from information-seeking research and the field study reported in chapter 5. The constraint based framework was used to illustrate the occurrence and effects of two different kinds of information need: a broad-based need resulting in opportunistic search behaviour, and a more specific fact-checking need resulting in goal-driven search. In each case, it was possible to show how information-seeking is driven by the wider task, and also how found information changes or reveals the shape of the wider task as understood by the information seeker/writer.

In terms of the problem space, a broad-based need arose when the information seeker had no candidate solutions to meet her wider goal. At this point she was unable to say what information would trigger a solution proposal and hence was in a state of uncertainty. Information-seeking at this point can at best be exploratory. When she found information to trigger an idea, this was represented as a new constraint—a relatively vague intention or focus defining a solution sub-class, and hence reshaping the problem space. Information-seeking then focused on testing the class of solutions against other constraints to assess its viability. In the second example, an emerging potential conflict with the *accuracy* constraint and a solution proposal incorporating a proposition *p* led to fact checking. This information-seeking was understood as clarifying the boundary of the accuracy constraint with respect to the solution proposal.

The ideas presented in this chapter have been concerned with writing with embedded information-seeking in general, and link information-seeking with design through writing as design. The design orientation focuses on the user's task, and yet it is broadly *cognitive* insofar as the problem space is considered from the users perspective—*i.e.* the problematic situation. However it is also

situated in as much as it accounts for dynamic responses to an unfolding situation. An analogy that might be useful is to think of information relating to the writing process in the same way that building materials relate to architecture. The discovered properties of building materials and the consequences that these have for how materials can be combined into a constraint optimising solution impact on the decisions of the architect. Similarly, the information seeker/writer will be influenced by the properties of the information he or she finds and the consequences that these have for how they can be combined into a coherent argument.

This chapter began with the question, 'what is the nature of the information seeker/writer's wider task?' And this was interpreted as the question of how writing tasks can be represented and understood in a way that makes known information-seeking phenomena meaningful. By framing writing as a design problem, ideas from the psychology of design have been used to develop an account which easily integrates embedded information-seeking and related phenomena. The design perspective represents the writer/information seeker's problematic situation as a constraint space which evolves structurally through exploration and experimentation, creative insight, and the making, reviewing and adjusting of commitments. On this view, uncertainty is a natural part of addressing an unstructured problem space. An implication of the design perspective is that to understand a user's problem, one needs to understand the constraints as construed by them, and understand that these change from moment to moment.

In the following chapter, the model of journalists' information behaviours reported in chapter 4 forms the basis for a discussion of requirements for integrated information-seeking and authoring systems. A prototype system design which seeks to address a sub-set of these requirements is also described.

Chapter 6

Requirements and design

6.1 Introduction

In chapter 4 of this thesis, an empirically derived model was reported describing the information behaviour of national newspaper journalists engaged in writing news reports and feature articles. An important part of the rationale for the model was to inform design requirements for journalist's information systems. The model was concerned with information behaviour in relation to ENC archives, *i.e.* not simply what information is sought and how, but also how it may be gathered, manipulated and put to use. This scope was motivated in part by the view that the full range of information behaviour has been under-investigated (Wilson, 1999, Kuhlthau & Tama, 2001), and in part by the view that in the context of research and writing tasks, the seeking and manipulation of information and writing do not form separate, unrelated behaviours, but form a single, continuous flow of activity. Consequently, systems that optimally support this continuity would integrate multiple tools within a single system.

The current chapter has two aims. The first is to derive a set of design requirements for integrated information-seeking and authoring systems based on the model. The second is to describe a prototype system called NewsHarvester. NewsHarvester is a 'proof of concept' information retrieval and authoring system which was designed by the author to address a sub-set of the requirements, and was developed in collaboration with Microsoft Research Ltd., Cambridge, England. An evaluation of this system will be reported in chapter 7.

In section 6.2, the approach taken to defining requirements is described through a general discussion of the problem of requirements either over-specifying or under-specifying design. Then, in anticipation of subsequent discussions concerning how the requirements might be technologically achieved, section 6.3 briefly summarises some electronic information-seeking technologies including: full-text keyword information retrieval, latent semantic indexing, probabilistic information retrieval and information extraction. In section 6.4, the requirements themselves are described in a discussion which is structured in terms of the information behaviours that were captured by the model. In section 6.5, the NewsHarvester prototype is described.

6.2 An approach to requirements and their relation to design solutions

Sommerville and Sawyer define requirements thus:

> They are descriptions of how the system should behave, application domain information, constraints on the systems operation, or specifications of a system property or attribute.
> **(Kotonya & Sommerville, 1998, p.6)**

Within software engineering, a distinction is conventionally drawn between system requirements and design solutions. Collectively, requirements specify the constraints that a system should meet. A design solution, on the other hand, is one of potentially many possibilities which may (or may not) meet a set of requirements. Hence, the relationship between requirements and design solutions is one-to-many. Design requirements are more abstract than design solutions, but requirements can nevertheless be expressed at different levels of specificity. For example, a requirement might simply say that a system should support a particular kind of activity, or it might additionally give details of *how* that activity should be supported. There is no definitive level of abstraction for a requirement.

So, how specific should a good requirement be? By being too specific (*e.g. a hyperlink which launches the help facility marked 'help' should be located on the main toolbar just below the company logo*) there is a danger that a requirement will unnecessarily bias the designer towards one particular kind of design alternative. In defining requirements there is always a danger that an analyst may visualise a particular kind of design alternative and use this as a starting point for specifying requirements, rather than using the uninterpreted needs of the situation. A requirement should allow the designer maximum freedom to determine how it will be met, and so should be as abstract as possible.

However, in the other extreme, a requirement can conceivably be so abstract and vague that it provides little or no guidance for the designer (*e.g. the help facility should be easy to use*). Knowledge about a situation of use should inform

requirements as much as possible, and the more the analyst knows about this situation the more specific they are able to be. Requirements should always be specific enough to rule out the possibility of bad design, given what the analyst understands about the intended situation of use.

Related to this is the extent to which a requirement makes a commitment to, or assumptions about, the use of particular technologies. A set of requirements should avoid technological assumptions and commitments as far as possible, and yet they must also be technologically attainable. Consequently, the requirements analyst must have some idea about how they might be achieved. In this chapter, the importance of technological attainability is reflected through outline discussions of how each requirement might be achieved.

And so a balance must be struck. The challenge for the analyst is to define requirements that delineate no more *and* no less than what is understood as the entire space of achievable, good solutions. They should be as specific as knowledge of the situation of use and technological possibility permits and no more. The approach taken in this thesis will be to view requirement specificity as providing focus for design, whilst *unwarranted* specificity is undesirable and over-constraining.

6.3 A brief look at electronic information-seeking technologies

6.3.1 Information retrieval

The purpose of an information retrieval (IR) system is to indicate documents within a collection that match or satisfy a submitted request. In addition, a system may be able to deliver the documents themselves. The user's request is expressed in the form of query which acts as a specification for the documents of interest. On submitting this request, the user is shown a list of matching documents from which they can select individual items. When an item is selected, the system typically provides information necessary for obtaining the document, or if full-texts are stored electronically, delivers the document itself.

The match between queries and documents is achieved by an IR system by consulting an index which stores information about the association between terms and documents. One of the principal distinctions between different kinds of IR system relates to the ways in which documents are indexed. For example, an index for a bibliographic system might be created by the manual assignment of keywords to documents, the keywords being designed to capture dominant document topics. Indexes can also be created using words and/or phrases within titles, author's names or dates, and multiple indexes may work in concert using combinations of these.

6.3.2 Full-text keyword IR

Creating manual keyword indexes, however, is labour-intensive and, as digitally recorded collections become bigger, increasingly intractable. Also, inter-indexer agreement is typically poor (Bates, 1986a). Where full-text documents are available in electronic format, a common alternative is to automatically construct a keyword index using terms occurring in the documents. The relatively low set-up and maintenance costs of full-text keyword indexing makes this an attractive option and it is the principal method used, for example by Web and ENC archive search engines.

Generally speaking, full-text keyword IR systems match document terms against query terms. In other words, a document will only be matched against a query if it contains a term which appears in the query (or if a 'NOT' operator is used, if it explicitly does not contain such a term). The only caveat to this is that some pre-processing might be performed on documents and queries such as the removal of words which tend not to describe document topics ('stop-list' words such as articles, prepositions *etc.*) and the remaining words may be stemmed (suffixes removed). Retrieval is then based on matching content word stems. Sophisticated systems may also expand query terms using a thesaurus. Notwithstanding these manipulations, full-text keyword IR is broadly based on the idea of identifying query terms within documents. Accordingly, the user discriminates documents by specifying terms that they would expect them to contain. Many systems also extend the notion of query terms to include phrases as well as single words.

This relatively simple relationship between queries and documents can have advantages for users. For example, if the user wants to search the web for the address of a company which they know to be in Camden, London, they might include the words 'Camden' and 'London' in their query along with the company name, knowing that documents containing the address will also contain these terms and that documents that do not contain these terms are unlikely to be useful. Put more generally, a simple and relatively predictable relationship between queries and documents can, in some circumstances, allow users to exploit strategies to achieve high precision searches.

However, a problem with full-text keyword IR is that relevant documents that do not contain all the terms in the query (depending on the query logic), or even those that do not contain any of these terms, will be excluded from the results. Hence these systems place high demands on the user to formulate good queries. Where a user is looking for a way 'into' a new topic they may be unaware of many terms that would nevertheless useful for discriminating relevant from irrelevant documents. In these circumstances, full-text keyword IR can impose limits on recall. An alternative approach to full-text indexing which attempts to resolve this problem is Latent Semantic Indexing (LSI).

6.3.3 Latent Semantic Indexing (LSI)

Based on the assumption that documents with many words in common are semantically similar, LSI automatically generates an index which assimilates statistical associations between terms, determined on the basis of word co-occurrence across a collection of documents. The method creates an index by initially looking at the document collection as a whole (after stemming and the removal of stop-words) to see which terms occur in which documents. Each term is then assigned as an axis within a high-dimensional document-space in which each document is located according to term occurrences. This has the effect of placing similar documents in close proximity to each other.

Using a method known as *singular value decomposition,* this high dimensional space is then collapsed into far fewer dimensions by superimposing frequently co-occurring terms (assumed to be semantically related) over each other. Documents are then indexed according to the terms that occur within them *and* terms occurring within closely related documents. The result is that, even though

a document might not contain any of the terms in a query, it might nevertheless be considered relevant by the system by virtue of sharing content words in common with other documents that do.

In addition to helping to overcome the problem of potentially relevant documents being missed, LSI also has the advantage that documents, or extracts from documents, can be used as queries. Consequently, documents can be used to find similar documents, thus removing the need for explicit query construction altogether.

6.3.4 Probabilistic Information Retrieval

Probabilistic approaches to calculating document relevance with respect to queries have been somewhat influential in the field of information retrieval, and so the binary independence retrieval (BIR) model (Robertson, Sparck Jones, 1976), which provides the basic ideas on which these approaches are based, is reviewed here. Probabilistic Information Retrieval, however, was not considered as useful for operationalising the requirements derived in this chapter and so will not be discussed in this context. The reason for this is because the probabilistic approach rests upon the idea of ordering documents with the 'best' document first and of the user being prepared to give iterative feedback on the relevance of the top document. This ignores the importance of the date of publication dimension for document ordering in news archive results and perhaps assumes too much about journalists. The following account is drawn predominantly from Fuhr (1992).

The BIR model frames the IR problem as that of ranking documents in terms of the estimated probability of relevance with respect to a given query. An underlying assumption is that the difference between relevant and non-relevant documents can be characterised in terms of the distribution of terms within documents. Within the model, the occurrence or non-occurrence of each term (that occurs in the collection) within each document is represented in the form of a binary vector. Queries are represented as vectors in the same way.

A second assumption, the so-called 'independence assumption', corresponds with a first-order approximation that the probability of a document being relevant to a query is equivalent to the product of the probabilities of each individual term being relevant *i.e.* that each term independently contributes to document

relevance. Using a transformation according to Bayes' theorem, it is then possible to estimate the relevance of a document. Specifically:

$$\text{1. Estimated relevance of i}^{\text{th}}\text{ document} = \sum_{k \neq 1} \log\left(\frac{P_{ik}(1-Q_{ik})}{Q_{ik}(1-P_{ik})}\right)$$

Where P_{ik} is the probability of a term k occurring in the i^{th} document given that the document is relevant to a given query, and Q_{ik} is the probability of a term k occurring in a document given that the i^{th} document is not relevant to that query. Initial values of P_{ik} and Q_{ik} can be determined through user-trials over a subset of documents and queries. Using this data, when a query is submitted the BIR model uses equation 1 to generate an initial estimate of relevance that is used for ranking the collection of documents. A threshold can be used to decide how many documents to return to the user. Relevance feedback can then be used to provide additional information with which to adjust the estimates and so improve the document ranking.

Effectively, the process of probabilistic information retrieval is one of estimating the power that each term provides in making the discrimination between relevant and non-relevant documents. The method explicitly acknowledges uncertainty in making such judgements by using probability theory to quantify and adjust certainty.

6.3.5 Information Extraction

The goal of IR is to find relevant documents from collections in response to queries; for this reason information retrieval is sometimes referred to as 'document retrieval'. This contrasts with the aim of Information Extraction (IE) which is to find relevant information *in* documents in response to queries (Gaizauskas and Wilks, 1998). IR systems use word-occurrence data to characterise document content, which makes them particularly well-suited to topic-level requests, rather than requests concerning what documents might say about that topic. IE systems, on the other hand, use techniques drawn from computational linguistics in order resolve queries about what documents say.

In terms of linguistic analysis, IR systems operate at a purely lexical level. Text structure and its semantic implications are not a consideration in query matching. IE systems, on the other hand, take account of text structure in a way that, for example, can be sensitive to the distinction between 'man bites dog' and 'dog bites man'. An IE system might be designed to resolve queries about who did any biting and who was bitten. This is not to say that IE systems are superior to IR systems, but rather that they perform a different kind of function using a different kind of technology. The distinction between IR and IE technology can largely be understood in historical terms. IR systems have arisen out of Library and Information Science research in part as a means for automating card catalogue systems (Bush, 1945), whereas IE systems have originated out of research into rule-based systems in computational linguistics (Gaizauskas & Wilks, 1998).

IE systems employ a template filling strategy, according to which pre-defined sorts of information are extracted from free text and assigned to template slots. For example, a template for extracting management succession events might include elements that specify participating companies and individuals, the post involved and the vacancy reason[1] (Gaizauskas & Wilks, 1998). A system using such a template would scan documents looking for reports of management successions and, for each instance that was found, populate an instantiated template with canonical linguistic formulations of each of the specified elements. The extracted data would then be used to construct a database which could be used for applications such as data-mining and automatic summary generation, as well as more conventional querying.

IE systems are sophisticated, but they are limited insofar as the sorts of information that they extract (and hence the range of summary types which can be generated or queries performed) must be defined as an intrinsic part of system design. They cannot be constructed by the user on an *ad hoc* basis in the same way that a user might construct an IR query. Queries must be defined in advance, and so IE systems can only operate within highly constrained task domains.

[1] This task was part of the 1996 DARPA MUC-6 conference systems evaluation

IE systems employ natural language processing techniques to identify elements of information in text. However, recent IE research has seen a de-emphasising of complex linguistic theory in favour of pragmatically motivated and theoretically shallow approaches to text processing, on the view that deep language processing is unnecessary for the IE task (Gaizauskas & Wilks, 1998). For example, syntactic parsing based on a formal natural language grammar may be rejected in favor of simpler pattern-matching. The improvement that this has brought has been a reduction in processing time without performance loss (Gaizauskas & Wilks, 1998).

Full-text keyword information retrieval, latent semantic indexing and information extraction are three electronic information-seeking technologies which support user queries over free-text documents. They differ in their underlying technologies, the kinds of querying they support and the extent to which they must be tailored by-design for a particular purpose. Full-text keyword IR offers a deterministic and predictable input/output relation and is better suited to situations in which users are confident of terms that can discriminate relevant from irrelevant documents. LSI can appear more 'intelligent' than regular full-text keyword IR insofar as it is less dependent on direct query-document word matches. Full-text keyword IR and LSI match documents based on topic level specifications, whereas IE extracts more specific pre-defined information elements. IE consequently, is suited to well-formed, specific information needs. IE, however, is more domain-dependent by design.

In the next section requirements based on the field study reported in chapter 4 will be established supplemented by some discussion of the information-seeking technologies which might address them including the technologies discussed here.

6.4 Requirements

The field study reported in chapter 4 produced a model of journalists' information behaviours in the context of their constraints and resources. In this section, the information behaviours are further explored to consider requirements that can be inferred for integrated information-seeking and authoring systems. The taxonomy of information behaviours is reproduced in figure 6.1.

A straightforward approach to deriving requirements from the behaviours shown in figure 6.1 would be to address each individually on a case-by-case basis and to consider how each might ideally be individually supported. On this view, the user would be considered as performing activities independently from each other. However, in chapter 4 the idea was discussed that any single behaviour trajectory may not correspond with one, and only one activity. Activities may be combined—they may embed within each other in means-ends relations, or two or more activities may share a single broad behaviour trajectory in a marriage of convenience (*i.e.* polymotivation; Cole, 1996). For example, at the start of an assignment, a journalist will typically search an ENC archive to see whether a proposed angle is novel (exclusivity checking). But whilst doing this they might also keep an eye open for useful overview documents (seeking background overviews). By identifying one or two overview documents during this initial search, time might be saved later. The opportunity for the polymotivational combination of activities arises when multiple goals can share common behaviours.

3.1 Information-seeking
- 3.1.1 Exclusivity checking
- 3.1.2 Background information-seeking
  - 3.1.2.1 Seeking background overviews
- 3.1.3 Seeking evidence for a hypothesis
- 3.1.4 Information-seeking for feature comparison
  - 3.1.4.1 Seeking properties of past disasters
  - 3.1.4.2 Discovering/confirming what someone said
- 3.1.5 Confirming names and how to spell them
- 3.1.6 Identifying useful contacts

3.2 Information-gathering

3.3 Information reviewing
- 3.3.1 Reviewing information gathered during an assignment
- 3.3.2 Reviewing information read but not gathered during an assignment
- 3.3.3 Reviewing information read prior to an assignment

Figure 6.1 The taxonomy of information behaviors identified in the field study

There is a danger that activities considered independently might result in standalone functional units that can only operate independently from each other. These might unnecessarily restrict polymotivation, and consequently place limitations on flexible, efficient work. Ultimately, any usability advantage realised through the careful consideration of each activity in isolation might be cancelled-out by the new disadvantage that they must now be performed in isolation. And

so, an initial requirement, or perhaps meta-requirement, is that systems should support the sharing of common behaviour trajectories for related activities where appropriate. This requirement *is* somewhat abstract—what is meant by 'appropriate' and 'related' is left open. But it is expected that, as in the example above, information-seeking activities that share a common topic or aspect may provide many examples of 'related' activities.

1. The polymotivation requirement.
Systems should, where appropriate, support the sharing of broad behaviour trajectories for related activities.

Category 3.1 Information-seeking

Information-seeking is a general category with no specific requirements.

Category 3.1.1 Exclusivity checking

Once a news or features writer receives a new assignment from their editor, *exclusivity checking* is often their first task. *Exclusivity checking* is done to establish that a proposed report has not previously appeared in any national newspaper or magazine; local and foreign publications were beyond the scope of concern. The issue related to what readers of the newspaper were likely to have read previously.

In the context of exclusivity checking, similarity between reports is judged in terms of the adopted angle. Indeed, for a proposed assignment in its embryonic form, this is all that exists. Further, since the angle of a report tends to be expressed within the first sentence or two (according to the cut-from-bottom structural constraint), ENC archive search engines that display the first sentence or two of a document in their results listings were found to be particularly well suited for *exclusivity checking*. By reading the first sentence or two of a report, it is often possible to judge its angle without the need of viewing the entire document. Hence, time is saved.

Search engines that display the first sentence or two of each document in their results list are not uncommon, but this is not the only way in which document extracts can appear in results. Another possibility is to display one or two extracts of text in which the user's search terms appear—the so called 'key word in context' (KWIC) approach (Luhn, 1960). KWIC is used, for example, by

Google, Lycos, and AlltheWeb, and is a user-option in many LexisNexis products, including those commonly used by journalists.

There are good arguments for KWIC—first, it allows the user to see the context in which some or all of their search terms occur within a document. This can help with the elimination of irrelevant matches resulting from unanticipated homonyms[2]. Second, in the case of well formed fact-checking queries, KWIC summaries can sometimes reveal the answer to a user's information need. In an experimental comparison by Offer (2003), KWIC was preferred by users compared with the technique of showing the first few lines of a document. Checking an assignment angle for exclusivity, however, is a special case in which including the first sentence or two in results list summaries would clearly be useful. Of course, the two approaches are not mutually exclusive—they can be used in combination, but each will make its own demands on screen real-estate and result in greater need for scrolling.

A first requirement relating to *exclusivity checking,* then, is that results displays should include the first couple of lines of each document in results summaries. However, given the polymotivational requirement, and the possibility that a single results set might be used for multiple activities, greater flexibility would be achieved by allowing selection (and de-selection) of first sentence displays at the results screen. Hence:

2. Exclusivity checking - requirement 1.
Systems should include the option (selectable at the results display screen) of showing the first sentence or two of each document listed in search results.

For the journalists interviewed in the field study, the scope of publication for which *exclusivity checking* is relevant is limited to national newspapers or magazines, and since commercial ENC archives often store articles from many more sources, *exclusivity checking* could be performed most efficiently if either:

a) users were able to *limit* searches to national newspapers or magazines (resulting in smaller, higher precision results sets in relation to *exclusivity checking*);

[2] One of two or more words spelled and pronounced alike but different in meaning, such as the noun *quail,* a chicken-like bird, and the verb *quail, to* shrink back in fear.

or

b) users were able to tell the system to make the representation of such documents visually distinct to ease fast identification;

Again, given the polymotivational requirement, and the idea of permitting different but related activities to share single activity trajectories (in this case a single search), option (b) is preferred.

From this, a second *exclusivity checking* requirement can be stated—that systems should be able to prompt the user with documents from national newspapers or magazines. To reduce unnecessary display complexity, prompting should be user-selectable (and deselectable) at the results display.

3. Exclusivity checking - requirement 2.
Systems should prompt users (as an option selectable at the results display screen) with documents from national newspapers or magazines.

Category 3.1.2 Background information-seeking

In chapter 4 it was reported that the journalists who were interviewed often searched ENC archives for background information in order to provide depth to their reports, and as part of their preparation before conducting an interview.

Category 3.1.2.1 Seeking background overviews

In his study of Social Scientists, Ellis discussed the idea of "starter references" (Ellis, 1989, p.179). Starter references are documents which provide overviews or discuss key ideas and are suited to people new to a subject. Starter references can serve as a basis for further information-seeking perhaps via *chaining*. Ellis reported that documents that the social scientists found useful as starter references included review articles, collections of papers, bibliographies, abstracts, indexes and library subject catalogues. The journalists interviewed in the field study similarly reported that when they were looking for a "way in" to a new subject they search for background overview documents.

Ellis proposed that *starting* could be supported by IR systems that prompt the user with documents containing key ideas or overviews. Indeed, as discussed above, prompting at the results display better addresses the polymotivation requirement (*i.e.* there would be no-need to perform multiple searches) and so

this proposal is reiterated here. However, to reduce unnecessary display complexity this should be controllable as an option at the results display.

4. Seeking background overviews - requirement 1.
Systems should prompt users (as an option selectable at the results display screen) with documents likely to provide high background overview content.

This requirement presupposes a means for automatically identifying documents with high background overview content, and this is not a trivial problem. Ellis (1989) suggested exploiting citation patterns in academic papers to identify papers containing key ideas (frequently cited) and papers containing overviews (many citations). The absence of citations in cuttings precludes this (indicating that solutions may be domain specific), but the more general principle underlying Ellis' proposal is to exploit features indicating that a document is a good starter document within an identification heuristic, and this general approach might well be applicable.

For example, if systematic differences at the lexical level (*i.e.* vocabulary) were to exist between starter documents and non-starter documents, then an approach to this problem could be to use LSI to discover these differences (using a training set distributed across a broad set of subjects) and to then differentiate documents within a results set. Although such a characterisation would be specific to a collection, it would have the value of being data-driven with a general method that is genre independent.

The success of the LSI approach rests on the existence of systematic lexical differences between starter documents and non-starter documents. An alternative approach based on a human generated characterisation of starter documents, would be to intuit *ad hoc* methods for automatic differentiation. The journalists interviewed in the field study indicated that useful background overviews tended to be longer pieces, such as feature articles, backgrounders[3] appearing in broadsheets, often by known specialists and articles incorporating explicit chronologies. Further, where the subject matter on which an overview was wanted was a person, profiles or big interviews were sought. Similarly, profiles were sought for overviews of companies or organisations. These

3 A backgrounder is a piece which supports a main or 'lead' article in a newspaper by providing historical background information, often appearing in a sidebar beside the lead article.

characteristics are shown in table 6.1 analysed into dimensions and values along with some suggested, *ad hoc* identifying features.

Prompting the user with starter documents at the results display could be handled by a system translating the features in table 6.1 into metrics (categorical phenomena and those relying on background knowledge, such as whether a document is by a known specialist, may be less easy to automate in this way), and then either by calculating an aggregated score for each document, or by displaying each metric separately. Whilst maintaining separation between metrics may come at the cost of increased interface complexity, by aggregating them there is a danger that, from a user's perspective, the resulting variable might be overly abstract and opaque in its derivation and consequently lack meaning. Also, by maintaining separation, variations on tasks such as specifically looking for an interview would be supported.

Characteristic		***Suggested identifying features***
Dimension	*Value*	
Length	Long	Number of words
Source	Broadsheet	Source field
Historical content	High	Number of dates in text and/or scope between earliest date and latest date (or date of article).
Interview content	High	The proportion of direct speech (indicated by inverted commas) to all other text.
Explicit chronology content	High	Number of consecutive paragraphs or sections starting with dates, or the existence of a linear date order.
Author	Known specialist	Author (byline) field

Table 6.1 Characteristics suggested as identifying features of background overview documents on ENC services.

From this discussion, a revised requirement relating to seeking background overviews using the second method discussed can be established:

4. Seeking background overviews - requirement 1 (revised).
Systems should prompt users (as an option selectable at the results display screen) with documents

- above a user-defined length threshold
- originating from broadsheet newspapers
- above a user-defined historical content threshold
- within a user-defined historical scope
- above a user-defined interview content threshold
- above a user-defined chronology content threshold
- by a user-specified author(s)

Category 3.1.3. Seeking evidence for a hypothesis

In the field study, the journalists described using cuttings archives to find evidence for a hypothesis—usually the proposed angle of a piece. Usually they would 'skim' relevant cuttings for 'facts and figures'. Potential variation in the arguments being researched and the serendipitous nature of skimming mean that requirements are difficult to determine beyond the facilities normally provided by full-text information retrieval systems.

Category 3.1.4 Information-seeking for feature comparison

Information-seeking for feature comparison involves the determination of similarities and/or differences between two events of a similar type on some variable or dimension. The sub-sections in this category relate to specific types of comparison that were identified during the field study. Each of these will be explored.

Information-seeking for feature comparison very often occurs when a journalist is reporting an event and wishes to place it in a historical context by drawing comparisons with similar past events. In this case feature comparison forms part of background information-seeking. However, feature comparison might equally be performed as part of investigating a story angle.

Identifying information elements within reports of particular types of event is a problem for which information extraction (IE) is well suited. As discussed in section 6.3, though, IE system design requires that queries are determined in advance. In some cases, the dimensions on which comparisons are to be made are well defined, and can be articulated before reports of previous incidents are read and comparisons chanced upon. Consequently, an information system that can automatically identify specified information elements within particular kinds of reports would be valuable to users.

Since feature comparison queries relate to information elements within documents, rather than to the contents of documents as a whole, it may be useful to consider displaying results in some form of structured format (*e.g.* tabulated or graphical) in which presentation is restricted to the information elements being compared. For example, a user wishing to compare the turnout and victory margin of a recent election with previous years might be shown a table of the margin and turnout information. Such a representation, particularly

one detailing the information in canonical and therefore easily comparable forms would allow the user to shift attention between items relatively effortlessly. Such a display might also promote the visual recognition of patterns.

However, displaying information in isolation from its textual context may be misleading. Current IE technology is fallible (as indeed is IR) (Gaizauskas and Wilks, 1998) and cannot be relied upon with total confidence. Indeed, given the wealth of natural language forms in free text, error-free IE may be an unattainable goal. Further, performance errors with IE systems are potentially more problematic than with IR systems (Gaizauskas and Wilks, 1998, Cowie & Lehnart, 1996) not least because *real* intelligence can often be misattributed by users in the face of *apparent* intelligence (Weizenbaum, 1983).

A more prudent alternative would be to display the desired information elements embedded within some textual context (rather than as canonical forms). This would enable, and even enforce, interpretation of extracted information before comparative judgments were made. Hence, an information extraction equivalent of a KWIC display would be most appropriate. Further, as with the KWIC results display, the full-text source documents should be easy to obtain from this display, and, within the documents themselves, extracted elements should be clearly indicated. Hence the following requirements can be stated:

5. Information-seeking for feature comparison - requirement 1.
Systems should display information elements resulting from feature comparison searches as embedded within some surrounding source text content (as with KWIC displays).

6. Information-seeking for feature comparison - requirement 2.
Where information elements resulting from feature comparison searches are shown extracted from source documents, systems should be designed such that the source document for each element can be retrieved easily.

7. Information-seeking for feature comparison - requirement 3.
Systems should display source documents resulting from feature comparison searches with the extracted elements are clearly indicated within the text.

Category 3.1.4.1 Seeking properties of past disasters

Disaster reporting is a common newsroom activity and journalists frequently want to compare a disaster that they are reporting with previous similar events. The features that interviewees identified as frequently compared included *causes* (if known), *locations*, *casualty figures* and *emergency responses*. This suggests that IE technology could be used to extract this information from documents in response to a user-query.

However, a system which reports these features for *all* types of disaster for every search would have very low precision. There are many kinds of disaster (train crashes, airplane crashes, disease epidemics, terrorist attacks, earthquakes *etc.*) and on any one occasion a journalist will only be interested in one particular type. Further, the granularity at which the class of disasters of interest is defined may vary (*e.g.* train crashes, train crashes in Britain, train crashes in tunnels, *etc.*).

Given this variability, it may be useful to divide feature comparison searches into two components: a standard IR search which defines the document set of interest performed according to the user's specification, and then a feature comparison search over this document subset to display the pre-defined features. Not only would this give users the flexibility to define disaster types in any way they wish, it would also concur with the polymotivation requirement, since results of the initial IR search might also be used for other related information-seeking activities. From this, a requirement for information-seeking for feature comparison in general can be stated:

8. Information-seeking for feature comparison - requirement 4
Systems should be designed to perform feature comparison searches over existing, standard IR search results.

as can more specific requirements relating to the properties of past disasters:

9. Seeking properties of past disasters - requirement 1.
Systems should allow users to search for information elements that indicate the cause of a past disaster.

10. Seeking properties of past disasters - requirement 2.
Systems should allow users to search for information elements that indicate the location of past disasters.

11. Seeking properties of past disasters - requirement 3.
Systems should allow users to search for information elements that indicate the casualty figures of past disasters.

12. Seeking properties of past disasters - requirement 4.
Systems should allow users to search for information elements that indicate the emergency responses that were prompted by past disasters.

13. Discovering/confirming what someone said - requirement 1.
Systems should enable users to search specifically for reports of what specified people have said on a given subject matter.

14. Discovering/confirming what someone said - requirement 2.
Systems should enable users to search specifically for reports of what specified people have said given specifications of the gist of what was said.

Category 3.1.4.2 Discovering/confirming what someone said

Discovering or confirming what someone said was classified in chapter 4 as a feature comparison activity. Typically, it is performed in order to compare what was thought to have been said with what was actually said, or to compare something said recently with something said in the past. In the cases identified in the field study, users had in mind a specific speaker and the subject matter being discussed. In some cases the gist of what was believed to have been said was also known. This suggests the following requirements:

Category 3.1.5 Confirming names and how to spell them

In the field study, the need to locate proper name spellings was identified as a frequent activity. A reported strategy which makes use of standard IR systems was to perform a search for each of a series of plausible spelling variations, and to take the number of hits for each as indication of consensus and, based on this, accuracy. It was acknowledged by users, though, that this strategy is

fallible. One potential source of error would be if a particular spelling variation were to correspond to some other word thus biasing the results. A number of methods could be employed to mitigate against such error. First, and in common with the case of seeking properties of past disasters, a name search could be performed over an existing search result. This would effectively filter documents used for the name search, and would also accord with the polymotivation requirement. Second, IE technology could ensure that matching only occurred where a term is used in text as a proper name. This leads to the following requirements:

15. Confirming names and how to spell them - requirement 1.
Systems should be designed to perform name spelling confirmation searches.

16. Confirming names and how to spell them - requirement 2.
Name spelling confirmation searches should be performed over existing standard IR search results.

17. Confirming names and how to spell them - requirement 3.
Name spelling confirmation search functionality should match only against proper names in text.

Of course, such functionality would only be useful so long as a hit count is output as part of the search results. So:

18. Confirming names and how to spell them - requirement 4.
Name spelling confirmation search functionality should output the number of matches obtained for a spelling variation.

As well as the possibility for error, the journalists' strategy of performing multiple searches on a number of spelling variations is relatively time-consuming. The time required to perform a name spelling search could be reduced by allowing the user to input all anticipated spelling variations as a single query, performing the searches as a batch operation, and then reporting the frequency for each variation on a single results page. Hence the requirement can be stated:

19. Confirming names and how to spell them - requirement 5.
Name spelling confirmation search functionality should allow users to input multiple spelling variations which would then be individually matched as a batch process.

Category 3.1.6 Identifying useful contacts

In chapter 4, the need was discussed for the journalists to identify people who might be eligible to provide comment on an issue. In particular, some interviewees discussed finding 'experts' more or less opportunistically within a cutting they might be reading. On finding such comment they might typically consider the quality of what had been said in order to decide whether to contact them on some new, but related issue.

It is possible to imagine a more systematic and goal driven approach to this activity supported by a system that could automatically identify expert comments within texts. So,

20. Identifying useful contacts - requirement 1.
Systems should be designed to perform dedicated 'expert comment' searches.

It is conceivable to have an appropriately designed IE component capable of automatically identifying news report text that corresponds with expert commentaries. Of course, it would be essential to identify, not just any expert comment, but expert comment on a particular subject-matter; being able to comment on cancer research does not qualify someone to comment on computer hacking. This is rather like the case of identifying the causes, locations, casualty figures and emergency responses for past disasters, but wanting to do this only for a particular subset of disaster types. And similar to the disaster reporting case, the polymotivation requirement of combining behaviour trajectories for similar activities would best be served by dividing the search into two components: an initial standard IR search to filter documents for a particular subject matter, and a subsequent IE search to locate expert comment within the results—each initiated separately by the user. Hence, an expert comment search facility would best be included as an option on the results page.

21. Identifying useful contacts - requirement 2.
Expert comment searches should be performed over existing standard IR search results.

Category 3.2 Information-gathering

In the field study, interviewees described gathering information as they searched so that they could refer back to it later. Two methods for information-gathering were described: *printing and highlighting*, and Windows-supported *drag-and-drop* into a text editor. Printing was used when more time was available, when a large amount of information in a document was considered potentially useful, or when there was a requirement for mobility. The journalists described splitting the computer screen between the browser on which they were conducting their search, and a word processor 'holding document' in order to support their frequent drag-and-drop operations. Hence, the first requirement for information-gathering is that an integrated information-seeking and authoring system should feature this layout by design.

22. Information-gathering - requirement 1.
System interface layout should be such that the window for displaying source documents and a text editor window for gathering information are visible simultaneously.

And, of course, systems should support text drag-and-drop from the source document window to the text editor.

23. Information-gathering - requirement 2.
Systems should support textual drag-and-drop from the source document window to the text editor window.

Category 3.3 Information reviewing

Information review is a superordinate category with no specific requirements.

Category 3.3.1 Reviewing information gathered during an assignment

In chapter 4 it was reported that when writing, the journalists would frequently review the information they had gathered. During writing they would often 'weave' parts of it into their copy. When writing, the journalists would often work with their screen split vertically between the 'holding document' and the word-processor where the copy was being written. This arrangement is unsurprising given the description of writing as interspersed with frequent reviewing. Hence, a requirement for reviewing information gathered during an assignment is:

24. Reviewing information gathered during an assignment – requirement 1.
System interface layout should be such that a text editor window for gathering information and a text editor window for copy writing are visible simultaneously.

And, in order to support the easy transfer of information from the 'holding document' into the copy editor window:

25. Reviewing information gathered during an assignment - requirement 2.
Systems should support textual drag-and-drop from the 'holding document' window to the copy-writing window.

Category 3.3.2 Reviewing information read but not gathered during an assignment

In chapter 4 it was stressed that uncertainty often exists for the journalists during an assignment concerning what information will ultimately be useful, and hence what information they would gather into their 'holding document' or collection of printouts. Two reasons were reported for this, both of which relate to constraint uncertainty and change. First, journalists begin an assignment with only an outline idea of what they will ultimately write. As they progress through their task, these ideas are constantly shaped and reshaped (also observed in the exploratory study reported in chapter 3). Second, external influences often change the nature of the task itself—most often editorial staff will change the angle of the proposed piece. Because of this uncertainty, users may wish to revisit documents long after they were initially encountered.

Revisiting documents can be problematic in standard IR systems. Where a document has been printed it is usually easily available. If it is not, the user may browse through the list in the hope of recognising the document summary, if they know that a document is within a currently displayed set of results. If the results list from which the document was accessed is no longer available, then they may attempt to reconstruct a previous query and then browse. Although a browser might change the colour of a visited hyperlink, relocating a document in this way can be time-consuming and can fail.

An integrated information retrieval and authoring system could support the reviewing of source documents better by associating each copy-and-pasted

extract with a means through which its originating document could be easily re-displayed. From this a requirement can be stated:

26. Reviewing information read but not gathered during an assignment - requirement 1. Systems should maintain connections ('threads') between copy-and-pasted extracts and their source documents at the interface in a way that allows users to easily redisplay the original.

Where a user wants to revisit a document which has been read during an assignment (on screen), but from which no information has been gathered, there will be no extracts with which to associate redisplay functionality. In these cases, easy reviewing might be supported through a chronologically ordered display of read documents with integrated redisplay functionality. Browsers typically maintain document histories, but these are usually hidden from users and document titles are often 'opaque'. So in addition:

27. Reviewing information read but not gathered during an assignment - requirement 2. Systems should maintain a chronologically ordered display of read documents at the interface which supports easy re-display functionality.

Where a user does choose to perform a search to find a previously read document, it would be valuable to be able to perform the search on only those documents which have been displayed during an assignment as opposed to an entire collection. Given the vast difference between the number of documents that will be stored on a database and the small number that are read during an assignment, such a facility should make great improvements to search precision. So:

28. Reviewing information read but not gathered during an assignment - requirement 3. Systems should support full-text searching restricted to only those documents which have been displayed during the course of an assignment.

Category 3.3.3 Reviewing information read prior to an assignment

Occasionally, a journalist will wish to review a document which they read at some time prior to an assignment. Consequently, it would be useful if a facility for searching documents read during an assignment could be extended to

search documents read prior to an assignment. Although many of the news reports and feature articles that a journalist reads as part of their work and leisure may have been obtained from sources other than the ENC service at their workstation, an index of all documents which had been displayed by the system for each user would nevertheless enable much greater search precision for this kind of need. The final requirement is:

29. Reviewing information read prior to an assignment – requirement.
Systems should support full-text searching restricted to only those documents which have been at any time displayed to the current user.

6.4.1 Summary and discussion of the requirements

The full set of requirements is:

1. The polymotivation requirement. Systems should, where appropriate, support the sharing of broad behaviour trajectories for related activities.

2. Exclusivity checking - requirement 1. Systems should include the option (selectable at the results display screen) of showing the first sentence or two of each document listed in the search results.

3. Exclusivity checking - requirement 2. Systems should prompt users (as an option selectable at the results display screen) with documents from national newspapers or magazines.

4. Seeking background overviews - requirement 1 (revised). Systems should prompt users (as an option selectable at the results display screen) with documents

- above a user-defined length threshold
- originating from broadsheet newspapers
- above a user-defined historical content threshold
- within a user-defined historical scope
- above a user-defined interview content threshold
- above a user-defined chronology content threshold

by a user-specified author(s)

5. Information-seeking for feature comparison - requirement 1. Systems should display information elements resulting from feature comparison searches as embedded within some surrounding source text content (as with KWIC displays).

6. Information-seeking for feature comparison - requirement 2. Where information elements resulting from feature comparison searches are shown extracted from source documents, systems should be designed such that the source document for each element can be retrieved easily.

7. Information-seeking for feature comparison - requirement 3. Systems should display source documents resulting from feature comparison searches with the extracted elements are clearly indicated within the text.

8. Information-seeking for feature comparison - requirement 4. Systems should be designed to perform feature comparison searches over existing standard IR search results.

9. Seeking properties of past disasters - requirement 1. Systems should allow users to search for information elements indicating the causes of past disasters.

10. Seeking properties of past disasters - requirement 2. Systems should allow users to search for information elements indicating the locations of past disasters.

11. Seeking properties of past disasters - requirement 3. Systems should allow users to search for information elements indicating the casualty figures of past disasters.

12. Seeking properties of past disasters - requirement 4. Systems should allow users to search for information elements indicating the emergency responses that were prompted by past disasters.

13. Discovering/confirming what someone said - requirement 1. Systems should enable users to search specifically for reports of what specified people have said on a given subject matter.

14. Discovering/confirming what someone said - requirement 2. Systems should enable users to search specifically for reports of what specified people have said given specifications of the gist of what was said.

15. Confirming names and how to spell them - requirement 1. Systems should be designed to perform name spelling confirmation searches.

16. Confirming names and how to spell them - requirement 2. Name spelling confirmation searches should be performed over existing standard IR search results.

17. Confirming names and how to spell them - requirement 3. Name spelling confirmation search functionality should match only against proper names in text.

18. Confirming names and how to spell them - requirement 4. Name spelling confirmation search functionality should output the number of matches obtained for a spelling variation.

19. Confirming names and how to spell them - requirement 5. Name spelling confirmation search functionality should allow users to input multiple spelling variations which would then be individually matched as a batch process.

20. Identifying useful contacts - requirement 1. Systems should be designed to perform dedicated 'expert comment' searches.

21. Identifying useful contacts - requirement 2. Expert comment searches should be performed over existing standard IR search results.

22. Information-gathering - requirement 1. System interface layout should be such that the window for displaying source documents and a text editor window for gathering information are visible simultaneously.

23. Information-gathering - requirement 2. Systems should support textual drag-and-drop from the source document window to the text editor window.

24. Reviewing information gathered during an assignment – requirement 1. System interface layout should be such that a text editor window for gathering information and a text editor window for copy writing are visible simultaneously.

25. Reviewing information gathered during an assignment - requirement 2. Systems should support textual drag-and-drop from the 'holding document' window to the copy-writing window.

26. Reviewing information read but not gathered during an assignment - requirement 1. Systems should maintain connections ('threads') between copy-and-pasted extracts and their source documents at the interface in a way that allows users to easily redisplay the original.

27. Reviewing information read but not gathered during an assignment - requirement 2.
Systems should maintain a chronologically ordered display of read documents at the interface which supports easy re-display functionality.

28. Reviewing information read but not gathered during an assignment - requirement 3.
Systems should support full-text searching restricted to only those documents which have been displayed during the course of an assignment.

29. Reviewing information read prior to an assignment – requirement.
Systems should support full-text searching restricted to only those documents which have been at any time displayed to the current user.

In terms of information-seeking, the general model of interaction that was developed through the requirements was one of complementary *primary* and *secondary* searches. Primary searches provide initial filtering of documents based on subject matter and might typically return high recall searches. Secondary searches are performed over primary search results. They support additional search precision for those information needs that require the identification of particular kinds of document or parts of documents, such as seeking background overviews, seeking properties of past disasters, confirming name spellings and identifying expert comment. As a general approach, the use of primary and secondary searches should reduce the number of primary searches required. This would not only maximize the potential for polymotivation, but also allow users to gradually develop a better sense of orientation with the results set(s) they work with over the course of an assignment.

6.5 The Newsharvester prototype

6.5.1 Design Requirements

The purpose of this section is to outline the design of an experimental integrated electronic information-seeking and authoring system which was developed as a response to selected requirements, and which was subsequently evaluated experimentally (reported in chapter 7).

The requirements covered functionality to support a wide variety of activities including information-seeking, information-gathering and information reviewing. The aim of the requirements, in common with the behaviour model, was to take a holistic approach to the information behaviours within a specific task context.

However, greater focus was adopted for the system design and subsequent evaluation. The aim was to build a system which could enable a focused and controlled experimental user-evaluation (as described in chapter 7). It was considered that focusing on one requirement would better support this objective. The requirement that was used as the basis for development was:

This requirement was chosen in part for its potential generalisability, and in part because it addresses a wider aspect of information behaviour (*i.e.* beyond information-seeking) for which there has been little attention in the literature.

26. Reviewing information read but not gathered during an assignment - requirement 1.
Systems should maintain connections ('threads') between copy-and-pasted extracts and their source documents at the interface in a way that allows users to easily redisplay the original.

In the interests of supporting fluid interaction, the prototype was also built in accordance with the following requirements (not to be evaluated):

22. Information-gathering - requirement 1.
System interface layout should be such that the window for displaying source documents and a text editor window for gathering information are visible simultaneously.

23. Information-gathering - requirement 2.
Systems should support textual drag-and-drop from the source document window to the text editor window.

24. Reviewing information gathered during an assignment – requirement 1.
System interface layout should be such that a text editor window for gathering information and text editor window for copy writing are visible simultaneously.

25. Reviewing information gathered during an assignment - requirement 2.
Systems should support textual drag-and-drop from the 'holding document' window to the copy-writing window.

In the next section the design of a prototype system is described which was developed with particular reference to address a requirement which was regarded as relatively independent of task and domain: 26. reviewing information read but not gathered during an assignment – requirement 1.

6.5.2 Design

Figure 6.2 shows the NewsHarvester user-interface[4]. The system was designed in two stages. During an initial design phase Visual Basic was used to explore design solutions guided by the requirements. A number of design possibilities were explored during this phase and either kept or rejected. More will be said of the decisions made below. From this initial phase a single design was retained. This incorporated all the core functionality of NewsHarvester that is described here, but tended to be unstable. During a second phase, the design and functionality was replicated the using the C# programming language. This formed the basis of the versions used for the evaluation reported in the next chapter.

NewsHarvester is designed to allow the user to search a database of news reports, browse the results lists, and select and view full-text documents. Any extract from a viewed document can be dragged into an integrated text editor where it can be retained and optionally annotated, edited, or even incorporated into a new piece of writing. Central to the design is the feature that, when an extract is dragged into the text editor, the extract is automatically suffixed with a hyperlink (Autolink). When clicked, the hyperlink will navigate the document display to the document from which the extract was originally taken.

For the purposes of the evaluation (reported in the next chapter), NewsHarvester was connected to the Media News Archive search engine at the *European Journalism Centre* in the Maastricht. This search engine performs full-text Boolean keyword IR searches (with optional date parameters, word stemming and phrase matching) over a collection of media news articles (the Media News Digest) containing about 10,000 articles originating from newspaper and television sources. Each article in the collection is trimmed to about 200 words.

The NewsHarvester screen is shown as it would be when in use in figure 6.2.

[4] NewsHarvester was developed in collaboration with Mircrosoft Research, Cambridge, England.

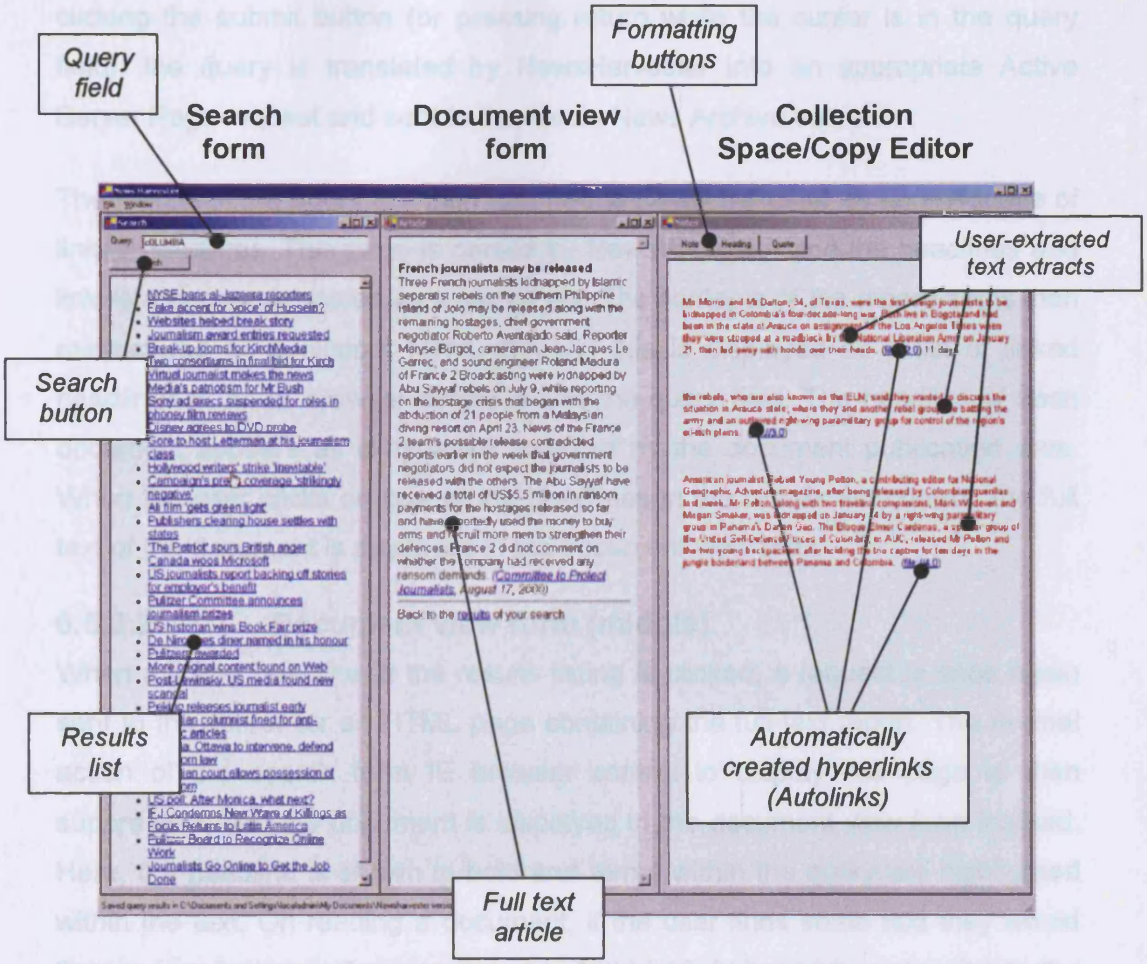


Figure 6.2 The NewsHarvester interface

The NewsHarvester interface is divided into three sections representing three independent but related work areas. In overview, the sections are:-

- **Search form (left)**
- **Document view form (centre)**
- **Collection space/copy editor (right)**

6.5.2.1 Search Form (left)

At the top of the search form there is a field in which the user can construct their queries. In the query language, a space between terms is interpreted as an AND operator. A term can be either a single word or a phrase. Phrases must be

entered within quotation marks. Only one phrase is possible in each query[5]. On clicking the submit button (or pressing return while the cursor is in the query field), the query is translated by NewsHarvester into an appropriate Active Server Page request and sent to the Media News Archive server.

The results of the query are then returned to NewsHarvester as an HTML file of linked headlines. This page is parsed by NewsHarvester and the headlines and link references are stored as a record set. The contents of the record set is then restructured as a custom HTML file and this is displayed as a list of linked headlines in an IE browser control below the query field. The headline of each document appears as a hyperlink followed by the document publication date. When the user clicks on one of the headlines in the results list control, the full text of the document is displayed in the *document view form*.

6.5.2.2 Document view form (middle)

When a linked headline in the results listing is clicked, a request is once again sent to the server for an HTML page containing the full-text report. The normal action of the search form IE browser control to display this page is then suppressed, and the document is displayed in the document view form instead. Here, the headline is shown in bold and terms within the query are highlighted within the text. On reading a document, if the user finds some text they would like to store for later reference, they can first highlight it, and then drag it into the *collection space/copy editor* using the mouse.

6.5.2.3 Collection Space/Copy Editor (right)

The *collection space/copy editor* acts as a repository for copied text, and also as a text editor in which the user can annotate extracts and write their finished report.

During the initial design stage of NewsHarvester three options for a collection space were explored. Initially, the system was designed so that each text extract that was dropped into a collection area would be automatically held within a visual object which could then be repositioned by the user with the mouse. This approach closely resembles the way in which searches are represented and manipulated, for example, in the SketchTrieve system (Hendry & Harper, 97). SketchTrieve allows users to search multiple resources from a single software

[5] Limitation imposed by *The Media News Archive* search engine

environment and to arrange and rearrange retrieval results (displayed as objects) on a virtual canvas or workspace.

This 'canvas' option for NewHarvester appeared to afford a great deal of flexibility in the visual structuring (and restructuring) of information. However, NewsHarvester was designed on the basis that it would be usable at a minimum screen resolution of 1152 x 862 pixels (higher than the average workstation), and since the collection space is only one of a number of work areas to be simultaneously displayed on the screen, it became clear that, the collection space would become cluttered very quickly. SketchTrieve overcomes this problem by using a 'virtual' workspace in which the work area extends beyond the limits of what is visible at any one time. This solution was also implemented for NewsHarvester, but given the limitations on the extent of the view of the workspace, orientation and navigation appeared awkward and potentially confusing.

The option that was ultimately chosen was to design the collection space as a text editor. This not only appeared to offer a more space-efficient way of storing text, but also allowed a navigation paradigm that would undoubtedly be familiar to an untrained user. The decision was also made to combine the collection space and copy editor into a single control, again to make the best use of a limited display area.

In the final system, when a text extract is dropped into the *collection space/copy editor*, it is distinguished by showing it indented and in a different font and text colour to typed text; and most importantly, it is automatically suffixed with a hyperlink. When the hyperlink is clicked, the *document view form* navigates to the document from which the extract was taken. The extract in both the document view form and collection space/copy editor are highlighted in yellow to indicate their association.

When an extract has been dropped into the Collection Space/Copy Editor and a link created, the extracted text can be edited by the user. If the user presses *return* while the cursor is within an extract, the extract is split into two separate paragraphs. When this happens, the link associated with that extract remains suffixed to the second part of the extract, and an additional link is added to the end of the first part.

The user can change the appearance of any text within the Collection Space/Copy Editor. Using the buttons at the top of the form, extracted text can be reformatted to appear as typed text and vice versa. Also using these buttons, any text can be formatted in bold to act as a sub-heading.

6.6 Discussion

This chapter has begun to address research question 3: the implications of the findings for integrated information retrieval and authoring systems for use by journalists (and others). The requirements set out in section 6.4 were motivated by the information behaviors discovered through the exploratory and field studies reported in chapters 3 and 4. As such, the requirements explicitly link the study findings with design implications.

Some of the behaviours identified in the empirical work may be fairly task or domain-specific, and where this is the case associated requirements are specific in the same way. Examples include the various kinds of feature comparisons that the journalists want to perform. These findings might not travel particularly far outside of the journalism task domain. Conversely, where behaviours do generalise across tasks or domains, so will their associated requirements. Examples may include the idea of enabling full-text searching of only those documents which have been displayed during the course of an assignment. It is easy to imagine that this requirement might be useful in virtually all online research and writing tasks.

The requirements cover information-seeking, information-gathering and information review as described in that early empirical work, and as such adopt the same scope as the study findings. Undoubtedly there are many other requirements that could usefully contribute to the delineation of 'the space of achievable, good solutions' for integrated information-seeking and authoring systems for use by journalists (or a wider population).

Such requirements might correspond with a broadening of the information behaviour model or with elaborating its detail. For example, the model in chapter 4 featured some quite specific behaviours classified under some quite general headings. Only two types of feature comparison were described: comparing properties of disasters, and comparing things that people in the news had said.

Perhaps journalists would find it useful to be able to compare other features, such as company turnovers or book sales.

The idea of integrating information-seeking and authoring tools within a single system provides a new opportunity for interlocked system components such that behaviour in one component affects events in the other. A potential criticism can be made that ultimately only one of the requirements made use of this possibility. This was the idea that a system should maintain 'threads' between copy-and-pasted extracts and their source documents at the interface—the requirement that motivated the design of NewsHarvester.

The requirements detailed in this chapter are not presented as complete, but perhaps, rather as opening up opportunities for elaboration, exploration and validation. This latter aim is the one which is taken-up next. In the next chapter a study is reported in which the Newsharvester prototype, and aspects of the model on which its requirements were based, were experimentally evaluated.

Chapter 7

Experimental evaluation of aspects of the model and the prototype

7.1 Introduction

In chapter 6, a series of requirements for integrated electronic information-seeking and authoring systems for journalists were described. The requirements were based on the findings from the empirical work reported in chapters 3 and 4. Chapter 6 also described a prototype system called NewsHarvester, which was designed on the basis of a sub-set of these requirements. The current chapter reports a study performed with NewsHarvester that was intended to evaluate the Autolinks functionality and to evaluate and elaborate on its theoretical foundation. The study contributes further to research question 3.

NewsHarvester allows users to search a database of news reports, browse the results, and view full-text documents. Text extracts from full-text documents can be dragged by the user into an integrated text editor. There they can be retained, modified and incorporated into new text. With Autolinks, any text that is dragged into the text editor is automatically suffixed with a hyperlink, which when clicked, forces the document display to navigate to the document from which the extract was taken. This feature is a response to requirement 26:

> Systems should maintain connections ('threads') between copy-and-pasted extracts and their source documents at the interface in a way that allows users to easily redisplay the original.

The motivation for this requirement was the observation that when users seek and use documents as part of writing a news report, sometimes they want to review documents that they have read previously during the course of that task. This has been explained in terms of the writer's continually evolving idea of what they will write, and in terms of the potential for mid-assignment task changes; both of these resulting in changes to relevance or 'usefulness' judgements.

The evaluation described in this chapter was intended to serve two purposes. The first was to confirm and elaborate the theory underpinning the requirement. In this respect, the study was used to experimentally evaluate the claims that when researching and writing a news report, journalists do indeed want to refer

to source documents multiple times, and that they do this in order to relocate information that they hadn't originally identified as useful (*i.e.* relevant).

In the exploratory study, the need to re-consult documents occurred as a result of the subjects coming to realise that they wanted to include specific items of information that they had remembered reading but hadn't gathered. The field study also supported this view. However, it was considered during the design of the current study that there may be additional reasons for wanting to see previously read documents and that the study provided an opportunity for exploring these. For example, users may wish to review the context of information that has already been gathered, or simply to see whether there might be anything else worth including (without a specific piece of information in mind). Both of these goals are slightly different, and slightly less specific than wanting to relocate a particular piece of information.

The second purpose was to evaluate claims made for the Autolinks functionality within the context of a news reporting task (and thereby to provide evidence for the requirement). In this regard, the idea of dragging and dropping text with the automatic creation of Autolinks was compared with two traditional approaches for information-gathering and subsequent relocation. The two 'traditional' approaches were: standard dragging-and-dropping (without Autolinks), and printing documents with optional highlighting (using a highlighter pen). In order to perform a controlled comparison between Autolinks and these other approaches two additional versions of NewsHarvester were developed, each differing from the original only insofar as it provided one of the two comparator functionalities. This allowed a controlled comparison to be made with the only distinction between conditions being the means provided for gathering and relocating information.

Since it is likely that each of the approaches to be compared would carry implications for the user on a number of variables, a number of comparisons were made. Each comparison corresponded to a comparative claim made in relation to Autolinks. These included the ease of relocating previously seen documents as well as the ease of gathering information in the first place (in all cases 'ease' was interpreted in terms of user-cost). More generalised factors were also compared, such as the extent to which each functionality provided for flexible working and user enjoyment.

Throughout this chapter, the two kinds of claim that relate to the two broad aims of the study are distinguished as *theory* claims and *functionality* claims respectively. In section 7.2 the claims that were evaluated and their respective rationales are described in detail. In section 7.3 the study method is described. In section 7.4 the analysis and results are reported; and in section 7.5 the results are discussed.

7.2 The claims

The five *theory* claims that the study sought to evaluate are summarised in. Table 7.1.

Theory Claim 1	When researching and writing a news report, journalists often want to refer to a given archived news reports multiple times.
Theory Claim 2	Theory claim 1 can, in-part be explained by the emergent goal of relocating information that had not previously been identified as useful.
Theory Claim 3	Theory claims 1 and 2 can in-part be explained by the emergent goal of including specific items of information in a report.
Theory Claim 4	Theory claims 1 and 2 can in part be explained by the emergent goal of understanding the context of information which had previously been identified as useful.
Theory Claim 5	Theory claims 1 and 2 can in part be explained by the emergent goal of identifying additional information that could be included in a near-complete report.

Table 7.1 A summary of the five theory claims evaluated by the study

The five functionality claims that were evaluated are summarised in table 7.2. For clarity, each is expressed in terms of a user-interaction variable. For example, functionality claim 1 concerns *the cost to the user of gathering information.* In the table, the variable is followed by a relational statement which expresses how the *drag-and-drop with Autolinks* functionality was expected to compare with *printing with optional highlighting* and *standard drag-and-drop* in terms of the interaction variable. For example, functionality claim 1 says that the

user-cost of gathering information using drag-and-drop with Autolinks is lower than it is when using printing with optional highlighting, and also that the user-cost of gathering information with drag-and-drop with Autolinks is the same or less (*i.e.* no more) than with standard drag-and-drop. In this way, each functionality claim is composite. Each makes a claim about drag-and-drop with Autolinks in relation to printing with optional highlighting, and also a claim about drag-and-drop with Autolinks in relation to standard drag-and-drop. Each sub-claim was evaluated independently.

	Variable	***Relational statement***				
Functionality claim 1	The cost to the user of gathering information	Printing with optional highlighting	$>$	Drag-and-drop with Autolinks	$\leq$	Standard drag-and-drop
Functionality claim 2	The cost to the user of relocating gathered information	Printing with optional highlighting	$\geq$	Drag-and-drop with Autolinks	$\leq$	Standard drag-and-drop
Functionality claim 3	The cost to the user of relocating non-gathered information from documents that contain gathered information	Printing with optional highlighting	$\geq$	Drag-and-drop with Autolinks	$<$	Standard drag-and-drop
Functionality claim 4	The affordance of a dynamic and flexible way of researching and writing.	Printing with optional highlighting	$<$	Drag-and-drop with Autolinks	$>$	Standard drag-and-drop
Functionality claim 5	User enjoyment	Printing with optional highlighting	$<$	Drag-and-drop with Autolinks	$>$	Standard drag-and-drop

Table 7.2 A summary of the five functionality claims evaluated by the study

The rationale for making each of the relative claims was based on the principle requirement used in the design of NewsHarvester, observations made during the field study and beliefs about how NewsHarvester would relate to these. For example, the principle requirement for NewsHarvester was that it should allow the easily redisplay of documents from which the user had already gathered information. This is difficult with standard drag-and-drop and so it was important to validate that with Autolinking it would be easier. However, printing also makes this kind of relocation easy and it is not clear that it would be better with Autolinks. Consequently, the claim was made that drag-and-drop with Autolinks would better than standard drag-and-drop but *no worse* that printing with highlighting on the corresponding user-cost variable.

An example of where Autolinks might provide user-cost reduction over printing however, is with gathering information in the first place. In the field study, attention was drawn to the time it takes to print a document and how dragging-and-dropping would be preferred where time was short. Since gathering information using Autolinks requires the same user-action as standard drag-and-drop, it might be expected that here Autolinks would offer improvements over printing whilst being no worse than standard drag-and-drop. Hence the potential benefits of NewHarvester over competing methods of gathering information are complex and must be evaluated on the basis of a matrix of variables according to appropriate claims for the improvement or preservation of existing interaction properties.

7.2.1 Functionality claim 1

Functionality claim 1 states that the cost to the user of gathering information with drag-and-drop with Autolinks is less than with printing with optional highlighting, and equal to or less (*i.e.* no greater) than with standard drag-and-drop.

This can be re-expressed as the two sub-claims:

> **Functionality claim 1a:** The cost to the user of gathering information with drag-and-drop with Autolinks is less than with printing with optional highlighting.
>
> **Functionality claim 1b:** The cost to the user of gathering information with drag-and-drop with Autolinks is equal to or less (*i.e.* no greater) than with standard drag-and-drop.

For sub-claim 1a, determining the reference condition against which Autolinks is to be compared is not un-problematic. There are contextual factors to consider which will be highly variable in everyday life, such as how far away the printer is to be. To resolve this problem, it was decided to evaluate Autolinks against the best-case scenario for the reference functionality, *i.e.* a set-up which most favours the reference functionality. Demonstrating an advantage under these circumstances would make the best case for Autolinks. Hence, subjects were provided with a modern, dedicated printer placed close to their workstations in the printing with optional highlighting set-up.

Drag-and-drop with Autolinks uses the same operation for gathering text extracts as standard drag-and-drop, and so there can be a high expectation that the user-costs would also be the same. But to exclude the possibility of any unforeseen costs being introduced by the use of Autolinks, functionality claim 1b was included in the evaluation.

Notably, functionality claim 1b does not claim that the costs of gathering information using drag-and-drop with Autolinks are the same as with standard drag-and-drop, but rather it makes the claim that they are equal or less. This claim was used since the critical issue was not to establish that the experimental functionality is the same as standard drag-and-drop in terms of information-gathering costs, but that its user-costs are no greater. For example, it would be perfectly acceptable to find that its user-costs were lower. The goal is to establish the non-inferiority of Autolinks compared to standard drag-and-drop. This will have implications for the experimental and null hypotheses described later in the study.

7.2.2 Functionality claim 2

Functionality claim 2 says that the cost to the user of relocating gathered information when using drag-and-drop with Autolinks is equal to or less (*i.e.* no greater) than with printing with optional highlighting, and equal to or less (*i.e.* no greater) than with standard drag-and-drop.

This can be re-expressed as the two sub-claims:

> **Functionality claim 2a:** The cost to the user of relocating gathered information with drag-and-drop with Autolinks is equal to or less (*i.e.* no greater) than with printing with optional highlighting.
>
> **Functionality claim 2b:** The cost to the user of relocating gathered information with drag-and-drop with Autolinks is equal to or less (*i.e.* no greater) than with standard drag-and-drop.

For the user, relocating gathered information means finding a known item within the repository of information they have generated during the research and writing task. Where the repository has been generated by printing with optional

highlighting this will mean visually scanning printouts. Where it has been generated by drag-and-drop with Autolinks or standard drag-and-drop, this will mean visually scanning, and perhaps scrolling an on-screen text editor.

The Autolinks functionality was not designed with the goal of facilitating the relocation of gathered information better than more traditional methods. However, being able to find specific items of gathered information, and being able to find them with little time and effort, is arguably an important factor in promoting fluidity in the research and writing process. Indeed, it is the underlying motivation for gathering information in the first place. Consequently, a comparative assessment of these relocation costs was included in the evaluation. However, this was done without any particular expectation about the value of drag-and-drop with Autolinks compared with printing with optional highlighting or standard drag-and-drop. Hence, the functionality claim was formulated to test whether Autolinks incurs greater user-costs for relocating gathered information.

7.2.3 Functionality claim 3

Functionality claim 3 states that the cost to the user of relocating non-gathered information from documents that contain gathered information when using drag-and-drop with Autolinks is equal to or less (*i.e.* no greater) than with printing with optional highlighting, and lower than with standard drag-and-drop.

This can be re-expressed as the two sub-claims:

> **Functionality claim 3a:** The cost to the user of relocating non-gathered information from documents that contain gathered information with drag-and-drop with Autolinks is equal to or less (*i.e.* no greater) than with printing with optional highlighting.
>
> **Functionality claim 3b:** The cost to the user of relocating non-gathered information from documents that contain gathered information when using drag-and-drop with Autolinks is less than with standard drag-and-drop.

Whilst functionality claim 3 may appear convoluted, it is nevertheless the most important comparative claim under test, since supporting the easy relocation of

documents from which information has already been gathered is the principle requirement motivating drag-and-drop with Autolinks. It was therefore essential to demonstrate that it does this well. Theory claim 1 states that when researching and writing a news report, journalists often want to refer to a given archived news report multiple times. Theory claim 2 explains this in terms of the emergent goal of relocating information that has not previously been identified as useful, and Theory claims 3, 4, and 5 explain both of these with reference to the emergent goals of including specific items of information in a report, understanding the context of information which had previously been identified as useful, and identifying additional information that could be included in a near-complete report.

Central to these claims are the ideas that, when researching and writing a news story, a user's concept of what information is and isn't relevant or useful changes. Information considered non-relevant at one point in time might be considered relevant or useful later. Information the user wishes to relocate may or may not be in a document from which they have already gathered other information, but, where it is, drag-and-drop with Autolinks is intended to provide a method for fast document relocation by maintaining active links between previously gathered text extracts and their source documents.

7.2.4 Functionality claim 4

Functionality claim 4 states that gathering information using drag-and-drop with Autolinks affords the user a more dynamic and flexible way of researching and writing than gathering information by printing with optional highlighting or standard drag-and-drop.

This can be decomposed into the two sub-claims:

> **Functionality claim 4a:** Gathering information using drag-and-drop with Autolinks affords the user a more dynamic and flexible way of researching and writing than printing with optional highlighting.
>
> **Functionality claim 4b:** Gathering information using drag-and-drop with Autolinks affords the user a more dynamic and flexible way of researching and writing than standard drag-and-drop.

Functionality claims 1, 2, and 3 are concerned with the comparative ease with which drag-and-drop with Autolinks supports three different user-actions which themselves are components of research and writing tasks: gathering information, relocating gathered information, and relocating non-gathered information from documents that contain gathered information. Whilst these actions are significant to an evaluation of Autolinks, they form only part of the research and writing activity. The criticism could be made that, as a sole basis for evaluation, they offer an overly reductive framework. Whilst the experimental functionality might offer improvements over the reference functionalities for these particular actions, questions may remain concerning how it impacts on the task taken as a whole. For example, these actions could be so infrequent or unimportant that making them easier does not impact on the global activity in any noticeable way. Worse still, the experimental functionality might obstruct other user-actions in some unpredicted way.

To avoid this overly reductive bias, functionality claims 4 and 5 were included. These were designed to address global properties of the research and writing activity. Further, attributing any observed differences in these properties to differences between the functionalities under comparison was justified given the controlled design of the experiment, *i.e.* the *only* difference between conditions was the information-gathering and relocation functionalities.

The drag-and-drop with Autolinks functionality was designed to enable easy information-gathering, whilst being sympathetic to changes in relevance judgments and the need to revisit documents with ease. In terms of a holistic perspective, a key motivation can be said to be the design of a tool which accommodates a dynamic and flexible way of working. Hence, functionality claim 4 asserts that drag-and-drop with Autolinks affords the user a more dynamic and flexible way of researching and writing than printing with optional highlighting or standard drag-and-drop.

7.2.5 Functionality claim 5

Functionality claim 5 states that users enjoy researching and writing with a tool that supports information-gathering by drag-and-drop with Autolinks more than

they do with a tool that supports information-gathering by printing with optional highlighting or standard drag-and-drop.

This can be decomposed into the sub-claims:

> **Functionality claim 5a:** Users enjoy researching and writing with a tool that supports information-gathering using drag-and-drop with Autolinks, more than they do with a tool that supports information-gathering by printing with optional highlighting.

> **Functionality claim 5b:** Users enjoy researching and writing with a tool that supports information-gathering using drag-and-drop with Autolinks, more than they do with a tool that supports information-gathering by standard drag-and-drop.

Like Functionality claim 4, Functionality claim 5 is concerned with a holistic property of the user-system-interaction. Of all the functionality claims, it operates at greatest level of generality. Functionality claim 4 is concerned with supporting dynamic and flexible work. If it is assumed that users prefer working in a dynamic and flexible way, then Functionality claim 4 can also be taken as an assertion of user-acceptability. However, since it is possible to address user-acceptability more directly, this was done with Functionality claim 5.

7.3 Method

7.3.1 Design in brief

Three variations of NewsHarvester were developed, each implementing one of the information-gathering functionalities to be compared. These were: NHO (incorporating drag-and-drop with Autolinks), NHP (incorporating document printing functionality) and NHR (incorporating standard drag-and-drop). Each design variation corresponded to a condition in a three condition, repeated measures, cross-over design. The study was structured as an independent comparison of an experimental condition (NHO) against two reference conditions (NHP and NHR), according to a single set of variables.

In each condition, subjects were given one of three news-report writing tasks to perform. The tasks were independently validated as representative of assignments typically given to newspaper reporters. Condition sequences were counterbalanced using a 3-group, single Latin square design. Task sequences were the same for all subjects. The dependent measures were post-task questionnaire items and quantitative measures derived from task protocols.

7.3.2 The Three conditions

The design of the NHO version of NewsHarvester was as described in section 6.5.2. The interface is shown in figure 7.1 (repeated from figure 6.2). Features common to all three versions (conditions) include the search form (left), the document view form (centre), and the collection space/copy editor (right). In all conditions the system enabled subjects to conduct Boolean searches over the Media News Digest at the European Journalism Centre.

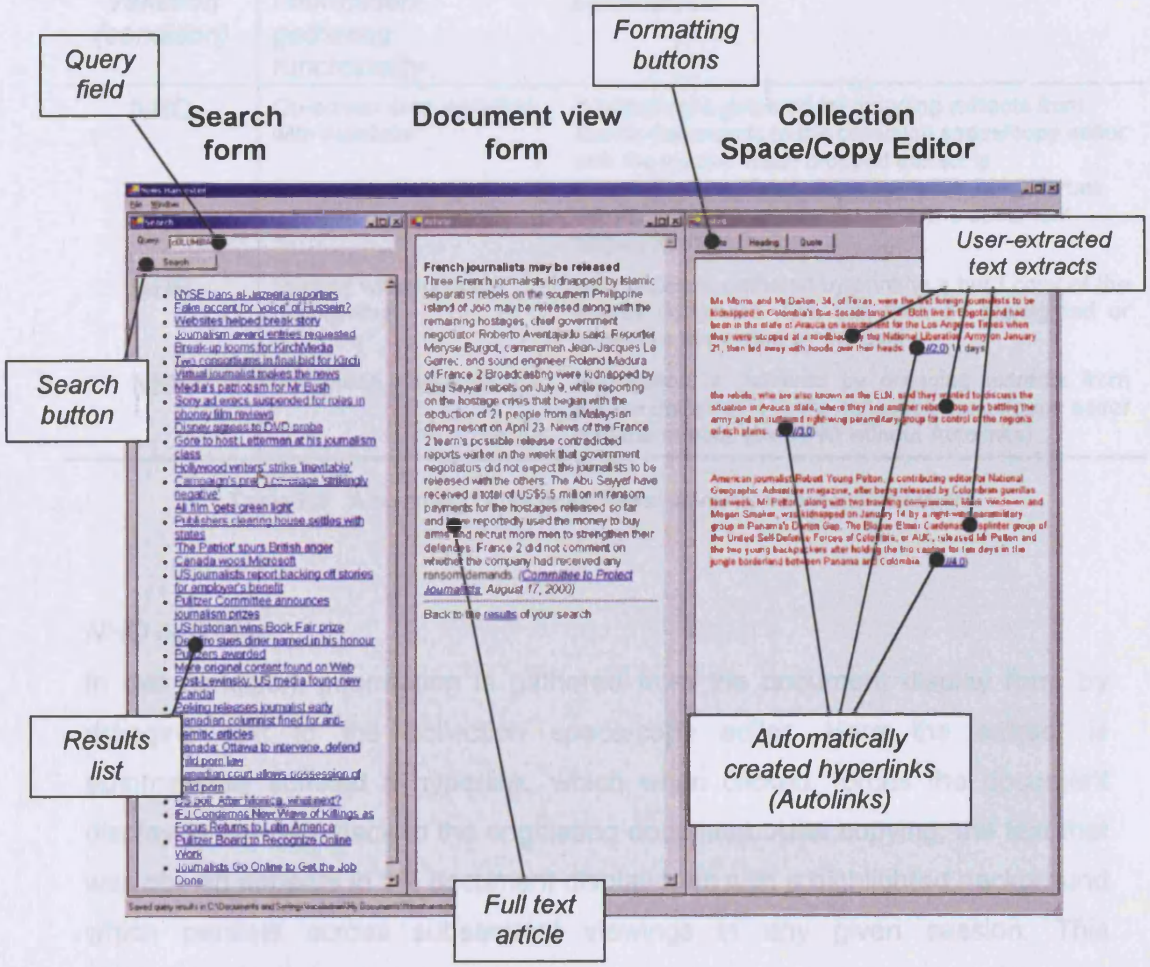


Figure 7.1 The NewsHarvester interface
(repeated from figure 6.2)

For the study, all versions of the system were configured such that, when a document was selected from the search form, the colour of the headline text would remain unchanged. The rationale for this was that this standard hyperlink feature offers the user a means for relocating previously read documents additional to those evaluated by the study. Consequently it's use might obscure any observable effects.

For each of the three conditions, a different variation of NewsHarvester was used, each implementing a different information-gathering functionality. In each case, NewsHarvester was run on a PIII 750MHz IBM laptop with 128MB RAM and fitted with an external keyboard, mouse and 19inch monitor. The screen resolution was set to 1152 x 864 pixels. The differences between variations (conditions) are summarised in table 7.3.

Variation (condition)	*Information-gathering functionality*	*Description*
NHO	On-screen drag-and-drop with Autolinks	Information is gathered by dragging extracts from source documents to the collection space/copy editor with the mouse. Each dropped extract is automatically suffixed with a hyperlink which forces the source document to display in the document display form.
NHP	Printing with optional highlighting	Information is gathered by printing a hard copy of the source document. This can then be highlighted or otherwise annotated by the user.
NHR	On-screen drag-and-drop	Information is gathered by dragging extracts from source documents to the collection space/copy editor with the mouse. (*i.e.* NHO without Autolinks)

Table 7.3 A summary of the three NewsHarvester variations/conditions

NHO condition

In this condition, information is gathered from the document display form by dragging text to the collection space/copy editor. Here the extract is automatically suffixed a hyperlink, which when clicked, forces the document display to navigate back to the originating document. After copying, the text that was copied appears in the document display form with a highlighted background which persists across subsequent viewings in any given session. This highlighting is yellow if the extract is the one taken most recently, or if it's auto-link is clicked; otherwise, it is coloured grey. Within the collection space/copy

editor, extracts are indented and appear in italicised, dark-red text. Text that is typed is not indented and appears as non-italicised, black text. Any of these text types can be re-formatted into the other, or formatted as a heading (large, black text) using buttons at the top of the copy/notes window.

NHP condition

In this condition, text could not be dragged and dropped into the collection space/copy editor. Instead, information could be gathered by printing the contents of the document display form. The top section of the NHP condition interface is shown in figure 7.2 showing the location of an additional print button. The printer was positioned on the desk, next to the computer. Pens were provided so that the user could highlight or otherwise annotate text in the printouts. Text typed into the collection space/copy editor appears as non-italicised, black text. Any text can be formatted as a heading (large, black text), or reformatted as normal text using buttons at the top of the copy/notes window.

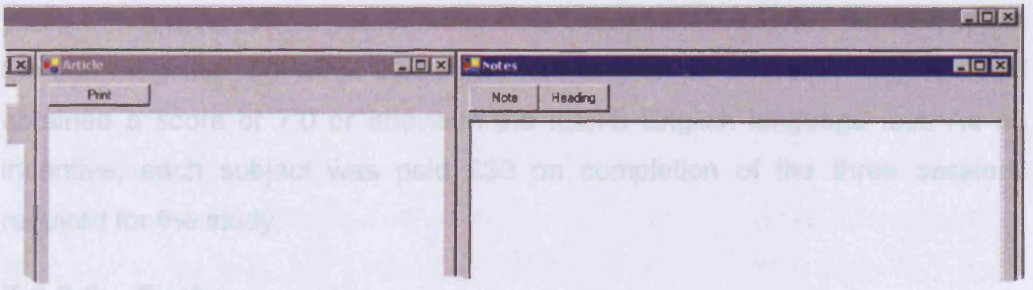


Figure 7.2 A view of the top-right-hand side of the NHP NewsHarvester interface

NHR condition

This condition is the same as the NHO version in all respects except that Autolinks do not appear when text is dragged into the collection space/copy editor. As with NHO, within the collection space/copy editor extracts appear indented, italicized and in dark-red text. Typed text is not indented and appears as non-italicised, black text. Any of these text types can be re-formatted into the other, or formatted as a heading (large, black text) using buttons at the top of the copy/notes window.

7.3.3 Subjects, tasks and counterbalanced conditions

7.3.3.1 Subjects

The subjects in the study were students recruited from the Department of Journalism at City University in London. Initially, 18 subjects were recruited, but one didn't complete the three sessions, and so, given the three-group counterbalanced design, 1 subject in each of the two remaining groups was randomly selected and their data discarded. Results are reported for the remaining 15.

13 of the remaining subjects were studying towards an M.A. in International Journalism and 2 were studying for a BA in Journalism. Many of the students studying on these courses, particularly the MA course, had previous professional experience in journalism. This made it possible to recruit only subjects who had worked previously as professional journalists. Of the 15 subjects used in the study, levels of experience ranged from 3 months to 10 years with a mean of 3.2 years. Many different nationalities were represented in the subject group, but all of those who had English as a foreign language had obtained a score of 7.0 or above in the IELTS English language test. As an incentive, each subject was paid £30 on completion of the three sessions required for the study.

7.3.3.2 Tasks

Three research and writing tasks were devised, one for each of the three conditions. The principle constraint in designing the tasks was that they should be typical of the assignments tackled by newspaper journalists on a daily basis. Informed by the field study reported in chapter 4, the following representativness criteria were devised:

- The tasks should require the reporting of a breaking news event.
- The tasks should be presented to subjects in the form of a newswire reporting the breaking story and a verbal editor's 'brief' defining the required angle and length of the news report.
- The tasks should necessitate background research into past news events.

- Approximately halfway through the task, the subject's attention should be drawn to a 'new development' accompanied by a revision to the required angle and word count of the original brief.

Further criteria were added in order to satisfy requirements of the experimental design. These were:

- The tasks should promote roughly equivalent amounts of background research.
- The tasks should be achievable within approximately 1 hour.
- The news event to be reported should be imaginary (to avoid the possibility of subjects finding a recent report from which they could copy wholesale) and yet credible in terms of the domain.
- The experimenter should run one pre-defined query at the start of the task (in order to mitigate against differences in search expertise). The initial query should be determined such that it returns a reasonably large results set that contains both potentially relevant and irrelevant documents.

The representativness criteria were determined on the basis of the newsroom study, but, given that the current study was intended to evaluate findings arising from that study, and also functionality designed in the light of those findings, it was important to obtain external verification of the representativeness of the tasks. For this reason, the tasks were reviewed by an Executive Editor at *The Times.* The review resulted in the recommendation that the newswires include more detail. This change was subsequently made.

For each task, subjects were initially handed a page of instructions and a newswire and given some editorial instructions in the form of a required angle and word count. 30 minutes into each task, the subjects were interrupted (unexpectedly for them) and handed a second newswire providing new information, and in the light of this new information they were given a new angle and word count. This experimental protocol was evaluated using a pilot subject (a City University PhD student). The pilot session validated the protocol as appropriate with no changes necessary. The newswires and instructions are shown in full in appendix IVa and IVb respectively.

Each condition was completed in a single session with no more than two sessions being performed on any one day.

7.3.3.3 Counterbalancing

To control for potential order effects between conditions (such as subjects becoming accustomed to a mid-task interruption) the order in which each subject performed the conditions was balanced using a single Latin square design. According to this design, each subject was randomly assigned to one of three sequence groups (shown in table 7.4). To balance for any unforeseen interactions between conditions and tasks (*i.e.* that some conditions might be more suited to some tasks), the task sequence was kept the same for all groups, so that each task was performed an equal number of times in each condition. Order affects between tasks were considered unlikely.

	Task 1	*Task 2*	*Task 3*
Group 1	NHP	NHR	NHO
Group 2	NHO	NHP	NHR
Group 3	NHR	NHO	NHP

Table 7.4 The three condition sequences

7.3.4 Metrics, hypotheses and hypothesis testing

The study used a combination of subjective and objective metrics (dependent variables) to assess the claims. For the subjective measures, subjects were asked to complete items on a post-task questionnaire at the end of each session. A single questionnaire was used for all the sessions, with different items being completed after each session (indicated by the experimenter), depending on the condition. Each questionnaire item consisted of a statement with an accompanying visual analogue scale (VAS) on which subjects could indicate their level of agreement with the statement. VAS scales were used in favour of Likert scales given their sensitivity to small differences. However, given the calibration difficulties associated with mid-range values on VAS scales, and also the importance for the study of between-condition differences (as opposed to absolute values), subjects were encouraged when rating an item to review any ratings they had previously given for the same item in relation to previous conditions. It was not uncommon for subjects to change previously given ratings

in order to adequately express a degree of relationship. The questionnaire is shown in appendix IVc.

One objective measure was used. A log of archive document consultations was made and from this the number of re-consultations was calculated. A document consultation was defined as a period of reading which could be interrupted by any other activity except for the consultation of another document. At such a point, the initial consultation was deemed to have ended and a new consultation begun. On the principle that the likelihood of an action correlates negatively with its expected cost, the frequency of re-consultations was taken as an inverse measure of user-cost, and between-condition frequency differences were used to test between-condition differences in reconsulation user-costs.

For the purposes of the study, a distinction is made between *claims* and *hypotheses*. The claims have been discussed at some length; however, it was the hypotheses that the experiment tested directly, these being statements of expected outcome in relation to the study metrics. For example, a hypothesis might state an expected between-conditions difference for responses to a questionnaire item. (The term 'hypothesis' is used here to reflect the conventional language of inferential statistical tests.) The relationship between the hypotheses and the claims is that the claims predicted the hypotheses. Consequently, tests on the hypotheses would support inferences about the accuracy of the claims. The hypotheses effectively contextualise the claims within the terms of the study. Since a given claim may have implications for more than one type of observation, it may relate to more than one hypothesis.

The claims evaluated by the study can be divided into two types: *superiority claims* and *non-inferiority claims.* A superiority claim proposes that the experimental functionality will be *better* than a reference functionality on some measure, whereas a non-inferiority claim proposes that the experimental functionality will be *no worse* than a reference functionality on some measure. For example, functionality claim 1a states that: **The cost to the user of gathering information with drag-and-drop with Autolinks is *less* than with printing with optional highlighting.** This is a *superiority* claim. Alternatively, functionality claim 1b states that: **The cost to the user of gathering information with drag-and-drop with Autolinks is equal to or less (*i.e.* no worse) than with standard drag-and-drop.** This is a non-inferiority claim.

This contrast similarly relates to the hypotheses that correspond with each claim. In the non-inferiority case, this has the consequence that conventional statistical difference testing is not a suitable methodology (Senn, 1997). The reason for this is that a non-inferiority hypothesis is shown by demonstrating that two cases are equivalent. In conventional difference testing, a hypothesis is accepted if it is shown that within a pre-defined confidence interval, two population means cannot possibly be the same, and so the null hypothesis is dismissed. But failing to show that two population means are different cannot be taken as a guarantee that they are the same (Senn, 1997).

The problem of statistically testing non-inferiority has been addressed in a type of drugs trial known as *an active control equivalence study* (Senn, 1997). Here, rather than comparing the new treatment against a placebo condition, the new treatment is compared against a reference condition to show that the new treatment is as good as the reference. In the current study the same logic can be applied to the non-inferiority claims. An experimental condition (drag-and-drop with auto-links) is compared with a reference condition (printing with optional highlighting or standard drag-and-drop) on some measure, in order to show that the new approach is not inferior.

A detailed explanation of the statistical method used in active control equivalence testing is given in appendix IVd. In summary, the approach is first to define a range of differences so small that they a negligible. This is known as *the region of practical equivalence* (Senn, 1997). Typically this is taken as ±20% of the mean of the reference sample. In this study, the boundaries were set slightly more rigidly at ±15% of the mean of the reference sample. Non-inferiority is then demonstrated by showing that the lower confidence limit for the difference between experimental and control population means lies above the lower bounds of the region of practical equivalence. If so, it can be concluded that the difference between the population means is above this level, and therefore that the experimental condition is at least as good or better than (*i.e.* not-inferior to) the reference condition.

7.4 Analysis and results

In this section, the reporting of the analysis and the results is organised in terms of the 5 theory claims and 5 functionality claims.

7.4.1 Theory Claim 1

> **Theory Claim 1:** When researching and writing a news report, journalists often want to refer to given archived news reports multiple times.

The calculated number of re-consultations for each session (irrespective of condition) was used to evaluate this claim. Given difficulties in quantifying the term 'often', a hypothesis was not formulated. However, out of the total of 45 assignments, there were only 5 which did not feature any re-consultations. The maximum number of re-consultations in a session was 15, and the mean per session was 6.18. Hence the data showed good evidence for theory claim 1.

7.4.2 Theory Claims 2, 3, 4, 5

> **Theory Claim 2**: Theory claim 1 can in-part be explained by the emergent goal of relocating information that had not previously been identified as useful.

> **Theory Claim 3:** Theory claims 1 and 2 can in-part be explained by the emergent goal of including specific items of information in a report.

> **Theory Claim 4:** Theory claims 1 and 2 can in part be explained by the emergent goal of understanding the context of information which had previously been identified as useful.

> **Theory Claim 5:** Theory claims 1 and 2 can in part be explained by the emergent goal of identifying additional information that could be included in a near-complete report.

Theory claims 2, 3, 4 and 5 were each evaluated using items on the post-task questionnaire. For each item, alternative and null hypotheses were based on the predictions of the respective claim. Since the VASs were calibrated at each end with "strongly disagree" (scored 0) and "strongly agree" (scored 10), a mid-point score of 5 was taken as signifying indifference. In each case, scores significantly above 5 were taken as agreement with the statement and support for the claim.

Subjects rated each statement once at the end of each of the three conditions. Since condition differences were not relevant for the theory claims, this provided three ratings per subject. For each subject, a mean rating was calculated across conditions. The decision to use parametric statistical tests was made based on normality testing of the distribution of the subject means using the Shapiro-Wilk test. In all cases normality was confirmed ($p>0.05$), although for theory claim 5 this was marginal.

For each statement, a one-tailed, one sample t-test was performed to test whether the means were distributed above a hypothetical sample mean of 5. In each case, the t-test was performed in consideration of the hypotheses:

Alternative Hypothesis (H_I):	VAS score (av.) > 5
Null hypothesis (H_O):	VAS score (av.) ≤ 5

Table 7.5 Shows the statements used to test each claim and summerises the results by showing the *p* value obtained from the normality tests, the mean VAS scores, and the *p* value obtained from each t-test.

The results show strong support for claims 2 to 5. The mean VAS responses were within the range 7.9 to 8.7 and in all cases were shown to be highly significant. Since the statements were so closely related, in particular, with statements associated with claims 3, 4 and 5 being concerned with a different explanation for re-consulting documents, two-tailed t-test comparisons were made between the results for each statement and each of the others. All of these test showed insignificant differences ($p>0.05$) demonstrating that no explanation was rated significantly higher than any other.

Theory claim	Statement	*p* (Shapiro -Wilk)	mean VAS score	*p* (t-test: H_1: VAS score (av.) > 5)
2	During the task, I found I wanted to re-consult source documents to find information I had remembered reading but did not necessarily consider useful at the time.	0.1527	7.928	<0.0001
3	This happened because later I found I wanted to include specific items of information in my report.	0.2402	8.470	<0.0001
4	This happened because later I wanted to better understand the context of information which I had identified as useful.	0.1732	7.988	<0.0001
5	This happened because later I wanted to re-consult source documents to check if there was anything else I could add to my report.	0.0686	8.638	<0.0001

Table 7.5 Statements and summarised results for claims 2, 3, 4 and 5

7.4.3 Functionality Claims 1a and 1b

1a) The cost to the user of gathering information with drag-and-drop with Autolinks is less than with printing with optional highlighting.

1b) The cost to the user of gathering information with drag-and-drop with Autolinks is equal to or less (*i.e.* no greater) than with standard drag-and-drop.

1a claims superiority of NHO over NHP on a measure of user-cost. 1b claims non-inferiority of NHO over NHR on the same measure. Evidence for this claim was sought by asking subjects to rate their agreement with the statement:

During the task, the actions I performed to ensure that I would be able to find useful information later, took very little time.

On this item, the raw VAS scores in condition NHP were normally distributed (Shapiro-Wilk, W=0.9241, p>0.05), whilst for NHO and NHR they were not (Shapiro-Wilk, NHO: W=0.7931, p<0.05, NHR: W=0.8149, p<0.05). Log transformations were performed using the function Log(p/1-p) (where p is the score expressed as a proportion of 1) (Senn, 1997). Since some subjects had rated the statement at 100% of the VAS, it was necessary to artificially extend the maximum and minimum points of the scale by 2% symmetrically about the

mid-point prior to log transformations. The transformed scores for conditions NHO and NHR *were* normally distributed (Shapiro-Wilk, NHO: W=0.9713, p>0.05, NHR: W=0.9101, p>0.05). However, transformed scores for NHP were not (Shapiro-Wilk: W=0.8691, p<0.05).

Given claims 1a and 1b the following hypothesis pairs were generated:

A	H_I: $u_{NHO} > u_{NHP}$	vs.	H_o: $u_{NHO} \leq u_{NHP}$
B	H_I: $u_{NHO} \geq u_{NHR}$	vs.	H_o: $u_{NHO} < u_{NHR}$

A Hypothesis pair **A** makes a comparison between the NHO and NHP scores, with the alternative hypothesis that NHO ratings will be higher than NHP. For both the raw and transformed scores, only one set of scores was normally distributed. A Wilcoxon test (one tailed) showed that the NHO scores were significantly higher than the NHP scores (W= 12, p= < 0.01). Hence functionality claim 1a is supported.

B Hypothesis pair **B** makes a comparison between NHO and NHR, with the alternative hypothesis that for NHO, ratings will be equal to or greater than those for NHP. The log transformed scores for NHO and NHR were both normally distributed, and so these were used in a parametric non-inferiority test. The results (summarised in table 7.6) show that the lower confidence limit is above the lower threshold of equivalence. Hence, H_o can be rejected in favour of H_I, and functionality claim 1b is supported.

$Mean_{(log\ transformed\ score)}$ NHO	0.79
$Mean_{(log\ transformed\ score)}$ NHR	0.521
Difference between means (NHO - NHR)	0.269
One-tailed 95% lower confidence limit (calculated using t statistic)	0.002
Lower threshold of equivalence (-15% mean NHR)	-0.078

Table 7.6 Claim 1b non-inferiority test summary (NHO-NHR)

By supporting H_I, the data shows that the NHO scores are either practically equal to the NHR scores, or that they are significantly higher. It is possible to

test the second of these possibilities. Since the lower confidence limit does not fall below zero, it can be concluded (at 95% certainty) that NHO is significantly higher than NHR. Hence, this data is also evidence for the stronger claim:

> 1b') The cost to the user of gathering information with drag-and-drop with Autolinks is lower than with standard drag-and-drop.

7.4.4 Functionality Claims 2a and 2b

> 2a) The cost to the user of relocating gathered information with drag-and-drop with Autolinks is equal to or less (*i.e.* no greater) than with printing with optional highlighting.

> 2b) The cost to the user of relocating gathered information with drag-and-drop with Autolinks is equal to or less (*i.e.* no greater) than with standard drag-and-drop.

In terms of the ease of re-locating gathered information, 2a and 2b claim that NHO is not inferior to either NHP or NHR respectively. Evidence was sought by asking subjects to rate their agreement with the statement:

> **During the task, the actions I performed to relocate this information took very little time.**

On this item, the raw scores for condition NHR and NHP were normally distributed (Shapiro-Wilk, NHR: $W=0.9013$, $p>0.05$, NHP: $W=0.9905$, $p>0.05$), although for NHO they were not (Shapiro-Wilk, $W=0.717$, $p<0.05$). This pattern remained for scores resulting from log transformations (following 2% extension of the scale about the mid-point) (Shapiro-Wilk, NHR: $W=0.9698$, $p>0.05$, NHP: $W=0.9818$, $p>0.05$, NHO: $W=0.8630$, $p<0.05$).

Given claims 2a and 2b, the following hypotheses were generated:

A	H_1: $u_{NHO} \geq u_{NHP}$	vs.	H_o: $u_{NHO} < u_{NHP}$
B	H_1: $u_{NHO} \geq u_{NHR}$	vs.	H_o: $u_{NHO} < u_{NHR}$

A Hypothesis pair **A** compares NHO and NHP scores with the alternative hypothesis that NHO ratings are equal to or greater than NHP. Both the raw scores and the log transformed scores were non-normally distributed, and so a nonparametric non-inferiority test was used based on the raw scores. The results (summarised in table 7.7) show that the lower confidence limit falls above the lower threshold equivalence. Hence H_o can be rejected and it can be concluded that the data supports functionality claim 2a.

$Median_{(raw\ score)}$ NHO	8
$Median_{(raw\ score)}$ NHP	5
Difference between means (NHO - NHP)	3
One-tailed 95% lower confidence limit (calculated using W statistic)	1.7
Lower threshold of equivalence (-15% median NHP)	-0.75

Table 7.7 Claim 2a non-inferiority test summary (NHO-NHP)

Further, and in common with functionality claim 1b, since the lower confidence limit does not fall below zero, it can also be concluded that the NHO scores are significantly higher than the NHP scores. Hence, the data is also evidence for the stronger claim:

> 2a') The cost to the user of relocating gathered information with drag-and-drop with Autolinks is lower than with printing with optional highlighting.

B Hypothesis pair **B** compares NHO and NHR scores with the alternative hypothesis that NHO scores will be equal to or greater than NHR. Both the raw NHO scores and the log-transformed NHO scores were non-normally distributed, and so a nonparametric non-inferiority test was performed on the raw scores. Table 7.8 summarises the results.

$Median_{(raw\ score)}$ NHO	8
$Median_{(raw\ score)}$ NHR	7.4
Difference between means (NHO - NHR)	0.6
One-tailed 95% lower confidence limit (calculated using W statistic)	0.2
Lower threshold of equivalence (-15% median NHP)	-1.11

Table 7.8 Claim 2b non-inferiority test summary (NHO-NHR)

Since the lower confidence limit falls above the lower threshold of equivalence, H_o can be rejected in favour of H_1, and hence the data supports claim 2b.

Again, since the lower confidence limit does not fall below zero, it can also be concluded (at 95% certainty) that the NHO scores are significantly higher than the NHP scores. Hence, the data is also evidence for the stronger claim:

> 2b') The costs to the user of relocating gathered information when using drag-and-drop with Autolinks are lower than with standard drag-and-drop.

7.4.5 Functionality Claims 3a and 3b

> 3a) The cost to the user of relocating non-gathered information from documents that contain gathered information with drag-and-drop with Autolinks is equal to or less (*i.e.* no greater) than with printing with optional highlighting.

> 3b) The cost to the user of relocating non-gathered information from documents that contain gathered information when using drag-and-drop with Autolinks is less than with standard drag-and-drop.

Functionality claims 3a and 3b were particularly important to the study since they address the principal issue that motivated the design of the drag-and-drop with Autolinks functionality, namely minimising the cost of relocating documents from which information had already been gathered. To reinforce the evaluation of these claims both subjective measures and objective measures were used.

Subjective measure

Subjects were asked to rate their agreement with:

> **Relocating information that I had not initially identified as useful, but which was in a document containing other information that I had identified as useful, took very little time.**

The raw scores for conditions NHP and NHR were normally distributed (Shapiro-Wilk, NHP: W=0.9437, p>0.05, NHR: W=0.9120, p>0.05), but the scores for NHO were not (Shapiro-Wilk, W=0.7777, p<0.05). The transformed scores were all normally distributed. (Shapiro-Wilk, NHO: W=0.9419, p>0.05, NHR: W=0.9422, p>0.05, NHP: W=0.9736, p>0.05)

Claims 3a and 3b led to the hypotheses:

A_{subj} H_I: $u_{NHO} \geq u_{NHP}$ vs. H_o: $u_{NHO} < u_{NHP}$

B_{subj} H_I: $u_{NHO} > u_{NHR}$ vs. H_o: $u_{NHO} \leq u_{NHR}$

A_{subj} compares NHO and NHP scores, with the alternative hypothesis that NHO scores are equal to or greater than NHP. Since both the NHO and NHP log transformed scores were normally distributed, a parametric non-inferiority test was applied to them. The results are summarised in table 7.9. Since the lower confidence limit is higher than the lower threshold of equivalence, H_o can be rejected in favour of H_I, and so the subjective data supports functionality claim 3a.

$Mean_{(log\ transformed\ score)}$ NHO	0.702
$Mean_{(log\ transformed\ score)}$ NHP	0.367
Difference between means (NHO – NHP)	0.336
One-tailed 95% lower confidence limit (calculated using t statistic)	0.046
Lower threshold of equivalence (-15% median NHP)	-0.1053

Table 7.9 Claim 3a non-inferiority test summary (subjective data, NHO-NHP)

Since the lower confidence limit also does not fall below zero, it can also be concluded that the NHO scores are significantly higher than the NHP scores. Hence, the data is also evidence for the stronger claim:

> 3a^I) The cost to the user of relocating non-gathered information from documents containing gathered information with drag-and-drop with Autolinks is lower than with printing with optional highlighting.

B_{subj} compares NHO and NHR scores, with the alternative hypothesis that NHO scores are higher than NHR scores. Since the log transformed scores for both

were normally distributed, they were used in a one-tailed, paired t-test. The test showed that the NHO scores *were* significantly higher than the NHR scores ($t=3.2$, $p<0.01$), and hence support claim 3b.

Objective measure

Given the assumption that the number of times a subject performs an action can be used as a measure of how easy they thought it was, a between-conditions comparison of the inverse frequency of an action is a comparison of its perceived user-cost.

In the NHO and NHR conditions, the only method for relocating non-gathered information was to redisplay the document. Indeed, this is the only reason why this might be done. In the NHP condition, relocating non-gathered information can be done by redisplaying the document onscreen *or* re-reading a printout (the latter being performed to relocate *either* non-gathered or gathered information).

In the NHO and NHR conditions, onscreen document consultations were logged. In the NHP condition both onscreen and off-screen document consultations were logged (in the latter case, where the text being accessed had not been highlighted). On the basis of these data, a count was made of the number of times documents containing previously gathered information were re-consulted for new information.

A precondition of re-consulting a document, of course, is that it should have been consulted in the first place. And since the number of documents consulted could vary from session to session, each session could vary in the number of re-consultation opportunities it presented. To avoid this confounding the data, the raw re-consultation counts for each session were used to calculate the mean number of re-consultations per document. These data were normally distributed for all conditions (Shapiro-Wilk, NHO: $W=0.9808$, $p>0.05$, NHR: $W=0.9382$, $p>0.05$, NHP, $W=0.9424$, $p>0.05$). The overall means for each condition were:

NHO: 0.601

NHR: 0.347

NHP: 0.473

Given claims 3a and 3b the following hypotheses were formulated:

A_{obj} H_I: $u_{NHO} \geq u_{NHP}$ vs. H_o: $u_{NHO} < u_{NHP}$

B_{obj} H_I: $u_{NHO} > u_{NHR}$ vs. H_o: $u_{NHO} \leq u_{NHR}$

A_{obj} compares NHO with NHP, with the alternative hypothesis that NHO is equal to or greater than NHP. A parametric non-inferiority test (summarised in table 7.10.) showed that the lower confidence limit was higher than the lower threshold of equivalence. Hence H_o can be rejected and it can be concluded that the objective data supports functionality claim 3a.

$Mean_{(raw\ score)}$ NHO	0.601
$Mean_{(raw\ score)}$ NHP	0.473
Difference between means (NHO - NHP)	0.129
One-tailed 95% lower confidence limit (calculated using t statistic)	-0.067
Lower threshold of equivalence (-15% mean NHP)	-0.0709

Table 7.10 Claim 3a non-inferiority test summary (objective data, NHO-NHP)

In contrast to the subjective data, the data did not provide additional support for the stronger claim $3a^{I}$.

B_{obj} compares NHO and NHR scores, with the alternative hypothesis that NHO scores are higher than NHR scores. Since the NHO and NHR average post-gathering re-consultation datasets were normally distributed, the hypotheses were tested using a one-tailed, paired t-test. This showed that the NHO scores were significantly higher than the NHR scores ($t = 3.44$, $p < 0.01$). Hence, the objective data supports claim 3b.

7.4.6 Functionality Claims 4a and 4b

4a) Gathering information using drag-and-drop with Autolinks affords the user a more dynamic and flexible way of researching and writing than printing with optional highlighting.

> 4b) Gathering information using drag-and-drop with Autolinks affords the user a more dynamic and flexible way of researching and writing than standard drag-and-drop.

Claims 4a and 4b propose that NHO is superior to both NHP and NHR in terms of promoting dynamic and flexible work. These claims were assessed by asking subjects to rate agreement with the statement:

> **The set-up I just used allowed me to work in what I regard as a flexible and dynamic way.**

On this item, the responses for all conditions were normally distributed (Shapiro-Wilk, NHO: W=0.9436, p>0.05, NHR: W=0.9183, p>0.05, NHP: W=0.9247, p>0.05). The mean responses were NHO: 8.213, NHR: 6.793 and NHP: 5.953.

Given claims 4a and 4b, the following hypotheses were tested:

A	H_I: $u_{NHO} > u_{NHP}$	vs.	H_o: $u_{NHO} \leq u_{NHP}$
B	H_I: $u_{NHO} > u_{NHR}$	vs.	H_o: $u_{NHO} \leq u_{NHR}$

A Hypothesis pair **A** compares NHO with NHP, with the alternative hypothesis that NHO ratings are greater than NHP. A one-tailed, paired t-test showed that NHO scores were significantly higher than the NHP scores (t=4.07, P<0.01), thus providing support for claim 4a.

B Hypothesis pair **B** compares NHO and NHR, with the alternative hypothesis that NHO will be greater than NHR. A one-tailed, paired t-test showed that NHO scores were significantly higher than the NHR scores (t=3.6, P<0.01), and so provided support for claim 4b.

7.4.7 Functionality Claims 5a and 5b

> 5a) Users enjoy researching and writing with a tool that supports information-gathering using drag-and-drop with Autolinks more than they do with a tool that supports information-gathering by printing with optional highlighting.

5b) Users enjoy researching and writing with a tool that supports information-gathering using drag-and-drop with Autolinks more than they do with a tool that supports information-gathering by standard drag-and-drop.

Functionality claims 5a and 5b assert that subjects enjoy using NHO more than they do NHP and NHR respectively. These claims were assessed by asking subjects to rate their agreement with the statement:

I enjoyed using the set-up.

On this item, the raw scores for all conditions were normally distributed (Shapiro-Wilk, NHO: W=0.8834, p>0.05, NHP: W=0.9705, p>0.05, NHR: W=0.9534, p>0.05). The means were NHO: 8.16, NHP: 6.56 NHR: 7.013.

From claims 5a and 5b, the following hypotheses were generated:

A) H_1: $u_{NHO} > u_{NHP}$ vs. H_o: $u_{NHO} \leq u_{NHP}$

B) H_1: $u_{NHO} > u_{NHR}$ vs. H_o: $u_{NHO} \leq u_{NHR}$

A A compares NHO and NHP scores, with the alternative hypothesis that, on this item, NHO will be greater than NHP. A one-tailed, paired t-test showed that NHO *was* significantly higher than NHP (t=2.98, p<0.01), thus supporting claim 5a.

B B compares NHO and NHR scores, with the alternative hypothesis that NHO will be greater than NHR. A one-tailed paired t-test showed that NHO *was* significantly higher than NHR (t=3.99, p<0.01). Thus claim 5b was supported.

7.4.8 Summary of results

All of the theory claims shown in table 7.1 were supported. Table 7.11 repeats the five functionality claims from table 7.1 and shows these in relation to the results that were obtained. In table 7.11, the system variations are referred to by the three letter codes introduced in section 7.3.1. The expected results are shown in black and obtained results are shown in red.

	Variable	*Relational statement*				
Functionality claim 1	The cost to the user of gathering information	**NHP**	$>$ $>$	**NHO**	$\leq$ $<$	**NHR**
Functionality claim 2	The cost to the user of relocating gathered information	**NHP**	$\geq$ $>$	**NHO**	$\leq$ $<$	**NHR**
Functionality claim 3	The cost to the user of relocating non-gathered information from documents that contain gathered information	**NHP**	$\geq$ $>_{subj}$ $\geq_{obj}$	**NHO**	$<$ $<_{subj}$ $<_{obj}$	**NHR**
Functionality claim 4	The affordance of a dynamic and flexible way of researching and writing.	**NHP**	$<$ $<$	**NHO**	$>$ $>$	**NHR**
Functionality claim 5	User enjoyment	**NHP**	$<$ $<$	**NHO**	$>$ $>$	**NHR**

Table 7.11 A summary of the five functionality claims (black) shown against the obtained results (red)

7.5 Discussion

The study reported in this chapter had two aims. The first was to evaluate the claim that when researching and writing a news report, journalists want to refer to some source documents multiple times, and to test a set of explanations for this (theory claims). The second aim was to evaluate a set of claims made for the Autolinks functionality about how well it supports gathering text and relocating information, and the extent to which it promoted dynamic and flexible working and user enjoyment in comparison with two traditional techniques for gathering and managing information (functionality claims).

7.5.1 Theory claims

The subjects used in the study all had professional experience in journalism and the tasks they were given were externally validated as representative of the kind of news writing task undertaken by newspaper journalists. This ecological validity included a mid-task interruption in which subjects were provided with new information and revised editorial instructions.

The results show that, when researching and writing a news report, journalists do indeed want to refer back to documents that they have read during an assignment multiple times. This happened in 91% of the assignments undertaken in the study. This result, though, is perhaps unsurprising, particularly given that the subjects' briefs were changed mid-task. Nevertheless, in the study, this question provided a context from which to explore the issue further.

The results robustly demonstrate that a reason for subjects wanted to refer back to previously seen documents was to find information that they had not considered useful when it had initially been read. In other words, their relevance judgments changed during the task. The results show that subjects wanted to re-consult source documents in order:

- to find specific pieces of information they had seen previously and later found they wanted to include in their reports;
- to better understand the context of information that they had already gathered;
- as part of a less focused review for additional information to add into their reports;

These findings add further support to the conclusion that was reached in the final section of chapter 4, that writing is an uncertain form of situated action, and of its plans evolving and changing in the face of the contingencies of a dynamic situation. As Suchman argued, it is often only through engaging in a situation that its possibilities become clear, and we do not know in detail the outcome of our activities when they begin (Suchman, 1987).

In relation to this, the findings also further corroborate the difficulty that researcher/writers can have in making definitive relevance judgments at the time when information is encountered. This feature of embedded information-seeking, which is key to the requirement motivating NewsHarvester's Autolinks facility, also corroborates the finding reported by Kuhlthau (1993), Kuhlthau and Tama (2001), Tang and Solomon (1998), Yang (1997), Vakkari (2001) and Spink *et al.* (2002) in terms of the systematic development of focus in respect of a wider task and the information seekers developing ability to make categorical relevance judgments.

7.5.2 Functionality claims

Autolinks offers a method for gathering and managing information which is 'sympathetic' to users' difficulty in making categorical relevance judgments. In the study, Autolinks was evaluated against dragging and dropping text into a text editor (without Autolinks), and printing with optional highlighting. The evaluation was made in terms of the user-costs imposed by gathering information, relocating gathered information, relocating non-gathered information from documents containing gathered information, and the extent that each approach affords a flexible and dynamic way of researching and writing and the extent to which users enjoyed using them.

Looking first at functionality claims 4 and 5, these are supported by the results—the subjects' responses show, with a high level of significance, that they enjoyed using the system that incorporated Autolinks more than the other two systems. The subjects also felt that Autolinks afforded a more flexible and dynamic way of working.

Functionality claim 3 said that with Autolinks the user-costs imposed by relocating non-gathered information from documents that contain gathered information would be less than with standard drag-and-drop, and no worse than with printing and highlighting. Whilst the results from the subjective and objective measures support this relationship with respect to standard drag-and-drop, they disagree in terms of the relationship with printing and highlighting. Whilst the objective measure suggests that the costs are the same with Autolinks as they are with printing and highlighting, the subjective measures support the stronger claim that relocating non-gathered information from documents that contain gathered information is *easier* using Autolinks. This discrepancy is considered in conjunction with stronger claims that were supported by the subjective measure results for functionality claims 1 and 2.

Functionality claim 2 said that the cost to the user of relocating information that had been gathered with Autolinks would be no more than it was with standard drag-and-drop or printing and highlighting. In fact, in both cases the data supported the stronger claim that relocating gathered information is easier with Autolinks. This result, however, seems implausible as a reflection of user-cost differences between Autolinks and standard drag-and-drop since the user-

actions involved are identical *i.e.* scrolling a text editor to find some text. The only possible explanation that sustains this results is that, in the Autolinks condition, users chose to gather less information as they searched and read documents in the knowledge that relocating them would be easier later. With less text in the collection space/copy editor, finding any one item of gathered information might be easier.

Finally, functionality claim 1 said that the cost to the user of gathering information with Autolinks would be less than it was with printing and highlighting and no more than with standard drag-and-drop. Whilst the subjective data supported this relationship with respect to printing and highlighting, it also supported the stronger claim that gathering information with Autolinks was easier than it was with standard drag-and-drop. Similar to the stronger claim that relocating gathered information is easier with Autolinks than with standard drag-and-drop, this result seems implausible since the user actions involved in gathering information in both the Autolinks and standard drag-and-drop conditions are the same.

Two further explanations are offered to account for these anomalies. The first is that the questions used in the post-task questionnaire could have been misinterpreted by subjects. The questions used to assess functionality claims 1 and 2 were (respectively):

> **During the task, the actions I performed to ensure that I would be able to find useful information later, took very little time.**

and...

> **During the task, the actions I performed to relocate this information took very little time.**

The first question was intended to relate to functionality claim 1 by asking about the time taken to gather information, i.e. dragging-and-dropping text or printing and highlighting documents (depending on condition). However, since subjects would have been aware that the Autolinks functionality was novel and the question was the first in the questionnaire (reproduced in appendix IVb) to ask about ease of use, this question may have been interpreted as referring to the

action of creating paths back to previously read documents, something for which the Autolinks functionality might be expected to receive a higher rating than standard drag-and-drop. Similarly, on the alternative interpretation, the second question (which addresses functionality claim 2) would also show a higher rating for Autolinks than standard drag-and-drop, because it would be taken as referring to the act of relocating previously seen documents, *i.e.* following the created paths.

The 'misinterpretation' explanation identifies potential question misinterpretation in the study that would explain the results obtained for functionality claims 1 and 2. Undoubtedly, the identification of these potential misinterpretations highlights shortcomings in the questionnaire design. However, this explanation does not explain the discrepancy between the subject and objective measures obtained for functionality claim 3. This discrepancy was that the subjective ratings suggested that Autolinks imposes a lower user-cost on the relocation of non-gathered information from documents that contain gathered information than printing and highlighting; whereas the objective measures suggested only that the costs with Autolinks were no greater.

A final explanation—and one that accounts for all of the obtained results—stems from the observation that in all cases the unexpected results showed a preference for Autolinks over a reference condition. The explanation is that the subjects' developed a general preference for Autolinks (demonstrated by the results obtained for functionality claims 4 and 5) and that this created a general bias in their questionnaire responses. Such a subjective bias has been recognised elsewhere and has been termed 'halo effect'. In the context of psychological studies, the halo effect has been described as, "A rater's failure to discriminate among conceptually independent aspects of a ratee's behaviour" (Saal et al, 1980, p.415). Essentially, a generally positive or negative orientation toward a ratee can influence their responses on all variables in the same direction. The halo effect interpretation appears to most fully explain the results obtained for the functionality claims.

In summary, whilst this study has shown good support for theory claims 1 to 5, the apparent bias observed in the measures for functionality claims 1 to 5 mean that the results cannot be accepted without qualification. In general terms, it can be argued that the results are supportive of the value of the Autolinks

functionality. If unexpected results were biased by a halo effect, this happened only because of the subjects developed a positive attitude towards using the Autolink functionality. That this is true is supported by their view that it was enjoyable to use and that it supported a dynamic and flexible way of working. The outcome was generally positive for Autolinks, although some questions remain, specifically in relation to functionality claims 1, 2 and 3. Whilst it is expected that they can be sustained, it has not been unquestionably demonstrated by this study. Nevertheless, it is claimed that the results demonstrate Autolinks as a credible solution to the requirement for integrated information retrieval and authoring systems to maintain connections between copy-and-pasted extracts and their source documents, and one that warrants further study and evaluation.

Chapter 8

Discussion

8.1 Introduction

Chapter 2 of this thesis provided a review of literature from Information Science, HCI and Cognitive Psychology which was intended to provide a perspective within which to locate the research questions and metatheoretical approach adopted, and to identify other research against which aspect of the thesis would later be referenced. Through the work of authors such as Paisley, Dervin, Belkin, Kuhlthau, Bates and Ellis in Information Science, and perspectives such as Situated Action, Distributed Cognition and Cognitive Systems Engineering in HCI, an increasing emphasis on the priority placed on understanding behaviour as it occurs within its natural context was noted as an emerging theme in both Information Science and HCI. The thesis adopted a similarly contextual and holistic orientation in attempting to understand the information behaviour of journalists and to consider the implications of this understanding for theory and for the design of integrated information retrieval and authoring systems.

In this, the final chapter of the thesis, the research that has been reported here is summarised with reference to the research questions that were set out in chapter 1. The chapter reviews how the research has addressed those questions and outlines the contributions that have been made. It also considers limitations of the work and discusses how future work might build on what has been reported.

8.2 Research question 1

Research question 1 asked:

What are newspaper journalists' prototypical information behaviours in relation to the seeking and use of information from electronic news cuttings services, whilst writing news reports and feature articles, *and* what are the aspects of their task situation that explains them?

The aim of this question was to establish a description of journalists' information behaviours in relation to ENC archive use as it occurs during news report and feature article writing. Motivated by the view that context is fundamental in

creating and conditioning work, the aim was to locate these accounts within an understanding of newspaper journalists' work context in order to explain their information behaviours and explain why they happen the way they happen. In other words, to characterize their information behaviours and to explain why these forms of behaviour are exhibited.

Chapter 3 began to address research question 1 through a lab-based exploratory study in which non-journalist subjects were asked to write a mock news report based on a fictitious newswire, and information that they would find by searching online resources including news cuttings services. The goal of the study was to identify areas of focus for a subsequent field study (*i.e. theoretical sampling*, Strauss & Corbin, 1990).

Screen recordings and talk-aloud protocols were analysed for global patterns of behaviour and for local behaviours. The global analysis revealed an initial phase of intense searching and note-taking followed by a clear-cut switch to writing, report reading and editing. But despite this initial period, searching could also occur during later stages of the task. Information needs, it was found, were occasionally deferred until later, apparently so that attention to a current activity could be maintained (*i.e. cognitive momentum*). The ordering of activities was also found to correspond broadly with a 'cascading and climbing' pattern. But despite being revealing about how information behaviour integrates within news writing activity, it was concluded that the global analysis contributed little towards design considerations for integrated retrieval and authoring systems.

Four local behaviours of interest were identified for which current ENC technologies appeared poorly suited. These were: biography seeking, quotation seeking, confirming proper name spellings and information-gathering. It was concluded that the four local behaviours may indeed be representative of what journalists do, or want to be able to do during news writing (with the minor caveat that *biography seeking* would be generalised to a need for overview documents), and also that the behaviours could be better supported through tailored system design. Consequently, these were taken forward into the subsequent field study. It was also observed in this study that although subjects were provided with tools to record the information they wanted to retain as it was encountered, when they came to write they wanted to relocate previously read source documents. This was also identified as an issue for further consideration.

Chapter 4 reported a field study carried out at *The Times* in London in which journalists were interviewed about their information behaviour, particularly in relation to the use of ENC services, during the process of news story and feature article writing. The results of this study were summarised in a model (shown in figure 4.3) of information behaviours framed within an explanatory context of news reporting constraints and resources. This model forms an important contribution of the thesis. Significantly, it extends beyond the traditional scope of information-seeking, to consider how information is manipulated and used as part of the task writing. Also, whilst being firmly grounded within the specifics of a particular constructive task, the basic elements and interactions represented in the model have the potential to generalise beyond this.

The model, which classifies behaviours under the major classes: *information-seeking*, *information-gathering* and *information review*, confirmed as representative and added significantly to the behaviours identified through the exploratory study. It was acknowledged in chapter 6, however, that the model featured some quite specific behaviours classified under some fairly general headings. Two types of feature comparison were described: comparing properties of disasters, and comparing things that people in the news had said. The potential set of types might be very large, but these two are considered representative. However, opportunity may exist for developing the leaf nodes of the model through further empirical study. Similarly, the top-level classification of: *information-seeking*, *information-gathering* and *information review* might not account for all information behaviours and may therefore be extendable through further empirical work.

The field study incorporated the idea from Rasmussen, Pejterson and Goodstein's (1994) Cognitive Systems Engineering approach that complex work is motivated, focused and guided by constraints and resources. Product constraints, such as the required angle and word count and the need to place events within a historical context, collectively represent the task goals, whilst resources, such as informants, ENC services, the 'holding document' and various kinds of knowledge are the tools providing the means by which the goals can be achieved. This framework proved a particularly useful perspective for understanding uncertainty and change in the journalist's work. Change, in the

model was represented through five types of dynamic influence between constraints, behaviours and resources and the wider environment:

1. **External influences on constraints**
2. **Influences of constraints on behaviour**
3. **Influences of resources on behaviour**
4. **Influences of behaviour on resources**
5. **Influences of resources on constraints**

Through evolving constraint determinants and resources, the model accommodates the idea of change in the user's perception of information relevance, or 'usefulness' and evolving information needs. As new information is encountered, as events unfold and as new insights are gained, so new "facts and issues" can become important. The model was presented as accounting for the finding from the exploratory study of users experiencing previously unanticipated information needs during writing.

8.3 Research question 2

How does this knowledge relate and contribute to more generalisable theory of information behaviour in relation to the processes and structure of complex information tasks?

Chapter 3 contributed to some extent to research question 2 by considering observations made in the exploratory study in the light of Suchman's Situated Action theory (Suchman, 1994). Some phenomena identified in the study demonstrated that the subjects' performance of the task was characteristic of Situated Action. Given the observation that, despite an initial preparation period, new information needs occurred once they had started writing, the subjects demonstrated both pre-planned and reactive behaviour characteristic of Situated Action. As they committed words to screen, so they found themselves with unanticipated information needs. It was concluded that Suchman's *vague plans* argument provided a useful perspective from which to understand these findings.

The model of journalists' information behaviour that was detailed in chapter 4 was both a summarisation of the findings of that study, and also a translation of those findings into a theoretical formulation. The thrust of the model as a theoretical formulation lies in the extent to which aspects of the model can be abstracted beyond the journalists interviewed and the specifics of the data which it seeks to draw together, to a more general account of task-motivated information behaviour in general. Whilst the specific information behaviours observed may or may not differ in other task domains, this is a separate question from whether the more abstract and structural properties of the model can generalise to other information task domains and provide a framework for understanding them. Specifically, these structural aspects include the ideas of:

- constraints (and their determinants) and resources providing an adequate account of information behaviour context;
- five kinds of interaction occurring between external influences, constraint determinants, information behaviour and resources;
- these interactions providing a basis for an understanding of uncertainty and evolution in information behaviour;

Testing the more general applicability of the model at this level of description represents a potentially fruitful opportunity for future work.

Chapter 4 also related the findings of the field study to existing theory by discussing them in the light of work by Ellis (1989), Bates (1989), Nicholas and Martin (1997), and once again returning to Suchman's Situated Action (1987). In terms of Bates' work it was noted that both the exploratory study and the field study support the Berrypicking model, but that they also add the observation that during the course of writing something, users may wish to 'pick' from a single document many times.

A comparison was also made between the field study findings and the information behaviour characteristics reported by Ellis (1989), and the purposes to which information is put by journalists as reported by Nicholas and Martin (1997). Despite some differences in scope and, in the case of Ellis' work, domain, the comparison showed that in many areas the models are mutually corroborative, and in others they extend upon one another. Finally, in chapter 4,

Situated Action was related to the idea of constraint determinant uncertainty and change within the field study model.

Chapter 5 broadened the focus of the thesis somewhat by generalising away from the specifics of journalistic writing to consider writing tasks in more general terms. It also changed the emphasis from a consideration of process to a theoretical consideration of the task at the heart of that process.

The point of departure for this chapter was the idea that writing is a class of design activity—an idea central to Sharples' (1996) model of writing as creative design. Chapter 5 identified and explored a number of parallels existing between findings from design psychology and those of information-seeking research in the context of complex task performance (including the field study). Four features of design problems and design problem-solving were discussed: incomplete specification; primary generators; the analysis/synthesis dynamic; and multiple constraints and integrated solutions. Each feature was exemplified by data from the field study.

Towards the end of chapter 5, the design perspective was used to motivate a constraint-based framework for representing writing tasks with embedded information-seeking. The framework forms the second major contribution of the thesis. The framework offered a situated explanatory framework for interpreting many information-seeking phenomena such as information-seeking uncertainty; the progressive refinement of information-seeking focus; and the reciprocal relationship between a user's evolving view of their task, the information that they find, and the information that they then want.

The framework was described through two example information needs. The first was of a broad information need arising from uncertainty about the task focus or primary generator. The second was a more focused, fact-checking need which arose from uncertainty about the relationship between a proposed solution and the boundary of a constraint. Accordingly, the framework can illustrate how information-seeking can shape the problem space in (at least) two ways: On the one hand, it can allow the information seeker to identify an opportunity, which, in turn, better defines their problem and so establishes new constraints. On the other hand, it can reveal the shape of existing constraints, and, in particular, how their boundaries correspond with different solution proposals.

Given that the framework has at its core the idea of representing a wider, productive task within which information behaviour is embedded, potential future work would include exploring its applicability to a range of information-seeking task domains, and also to a range of different kinds of information needs.

8.4 Research question 3

What are the implications of the findings for integrated information retrieval and authoring systems for use by journalists (and others)?

The aim of question 3 was to relate the empirical findings to both the design of information systems that integrate information retrieval services and text authoring tools by exploring and validating implications for design. To address this question, chapter 6 first used the model developed in chapter 4 as a basis for a series of design requirements for integrated information systems for journalists. These requirements represent the third contribution of the thesis.

Chapter 6 began by establishing an approach to requirements—that requirements should be as specific as knowledge of the situation of use and technological possibility permits, and no more. Requirement specificity provides focus for design, but *unwarranted* specificity is undesirable and over-constraining. Requirements must be technologically attainable, and so an outline was provided with each requirement for how it might be achieved.

Given their potential for generalisation to other domains and also the lack of attention given in the literature to information behaviour beyond information-seeking, some of the requirements that relate to information-gathering and review were used as the basis for the design of a prototype integrated information retrieval and authoring system called NewsHarvester. This system was described at the end of chapter 6 and is the final contribution of the thesis. As a product of the research, NewsHarvester offers a valuable exemplar which may be copied, adapted and further investigated, and may direct developers in designing operational systems

NewsHarvester incorporates functionality which was termed Autolinks. This was motivated by the requirement that systems should maintain connections ('threads') between copy-and-pasted extracts and their source documents at the interface in a way that allows users to easily redisplay the original. Autolinks is the final contribution made by the thesis. When an extract of text is dragged from an IR document display into NewHarvester's integrated text editor, the extract is automatically suffixed with a hyperlink. This link, when clicked, forces the IR document display to navigate back to the document from which the extract was taken.

In chapter 7, a study was reported which was designed with two purposes. First, it was designed to confirm and elaborate the theoretical claim underpinning the requirement that Autolinks was intended to address—that journalists do indeed want to review source documents in order to relocate information that they hadn't originally identified as useful. Second, it was designed to comparatively evaluate the Autolinks functionality in relation to two 'traditional' approaches for gathering information: standard dragging-and-dropping (without Autolinks), and printing documents with optional highlighting (using a highlighter pen).

The study showed that, when researching and writing news reports, users do indeed want to refer back to documents they have read previously during the assignment in order to find information that they had not initially considered useful. It also showed that this was in order that they might find specific pieces of information to include in their reports; to better understand the context of information that they had already gathered; and to find any additional information they might like to add to their reports. In this respect, the findings corroborated the idea that researcher/writers often have difficulty making definitive relevance judgments at the time when information is encountered.

Autolinks is a method for gathering and managing information in the light of this difficulty. The comparative evaluation between Autolinks, standard dragging-and-dropping (without Autolinks) and printing documents with optional highlighting, showed a strong user-preference on multiple measures for Autolinks over and above the other two methods, including the ease with which previously read documents could be relocated. Subjects rated Autolinks as imposing a lower user-cost than printing and highlighting or standard drag-and-drop for gathering information, relocating gathered information, and relocating

non-gathered information from documents containing gathered information. For this last action, inverse frequency data concurred the relationship with standard drag-and-drop, but showed user-cost as practically the same as for printing with optional highlighting.

In general, though, the subjective data appeared to unrealistically favour Autolinks. This was particularly striking where subjects rated the user-costs imposed by Autolinks more favorably than a reference functionality, but the user-action was apparently identical. It was concluded that subjects had felt positive towards Autolinks but had not always considered specifically what each question was asking *i.e.* the halo effect. This interpretation is supported by the significantly higher ratings given to Autolinks on the more general questions of promoting flexible and dynamic work, and user enjoyment. However, despite these anomalies, overall the results appear very encouraging for Autolinks.

Future development on research question 3 could include:

- Building and evaluating systems which address the remaining requirements.
- Further evaluation of the Autolinks idea, including in other task domains.
- Developing more sensitive evaluation metrics – focusing perhaps on objective measures.

8.5 Closing Remarks

This thesis began with an extract from Vannevar Bush's 1945 essay 'As We May Think'. Bush's essay depicted a future technology, which he called the memex, that would manage the storage of vast amounts of information, making this information instantly available to be linked and annotated by the user into new customised trails of information. Bush defined a compelling research agenda that is still alive today. He simultaneously envisioned PC workstations, information retrieval, hypertext and word processors. His essay has been cited as a key inspiration behind IR research, hypertext and the World Wide Web (Simpson *et al.,* 1996).

This thesis began and now ends with reference to Bush's memex. In their own right, each of the ideas incorporated into the memex has spawned distinct

research and commercial endeavors, and over the last half century enormous strides have been made towards transforming Bush's vision into reality. In the introduction to this thesis, though, it was emphasised that the memex was not just about information storage or information retrieval or hypertext or authoring in isolation. It was a vision of integrating a number of technologies within a tool that would embed within and enrich peoples' information tasks. This vision was taken from a perspective of what those tasks are, and how they are and might be performed. This is the viewpoint that has been taken by this thesis.

The specific contributions of the thesis have been:

1. A model of journalists' information behaviours in the context of the wider task of writing news reports and feature articles;
2. A general framework for representing writing tasks which accounts for a number of key information-seeking phenomena;
3. A set of design requirements for integrated information systems for journalists;
4. The design of a prototype system that is sympathetic to users' evolving interests as a function of their developing task focus;

More generally, though this thesis has engaged with, and hopefully contributed to, a broader vision. The perspective has placed center-stage the user "under constraints and pressures, creating products, drawing upon the elaborate communication network that connects him with sources of necessary knowledge" (Paisley, 1968, p2). Indeed, constraints and the elements of that communication network, or information resources, have been explicitly formulated in this thesis as research objects. The thesis has also explicitly engaged with the idea of an integrated information tool as proposed by Bush. NewHarvester, with its possibilities for gathering, annotating and authoring information and its ability to automatically retain document links without disrupting the flow of the user's work, exploits a number of technologies in much the way that the Bush intended, and arguably captures much of the memex idea. But NewsHarvester and Autolinks functionality was not designed under this premise, but rather under the premise that to design tools that will help people to do work, we must understand that work and the ecology within which that work naturally takes place.

References

Akin, O. (1986) **Psychology of architectural design**. London, Pion.

Anderson, R.J. Heath, C.C. Luff, P. & Moran, T.P. (1993) The social and the cognitive in human-computer interaction. **International Journal of Man-Machine Studies**, 38, pp. 999-1016.

Attfield, S., Blandford, A. & Dowell, J. (2003) Information seeking in the context of writing: a design psychology interpretation of the 'problematic situation'. **Journal of Documentation**, 59(4), 430-453.

Attfield, S. & Dowell, J. (2003) Information seeking and use by newspaper journalists. **Journal of Documentation.** 59(2), pp. 187-204.

Bates, M.J. (1986a) Subject access in online catalogs: A design model. **Journal of the American Society for Information Science**, 37, pp. 357-376.

Bates, M.J. (1986b) An exploratory paradigm for online information retrieval. IN: Brookes, B.C. ed. **Intelligent information systems for the information society**. Amsterdam: North-Holland.

Bates, M.J. (1989) The design of browsing and berrypicking techniques for the on-line search interface. **Online Review**, 13(5), pp. 407-431.

Belkin, N. (1978) Information concepts for information science. **Journal of Documentation**, 34(1), pp. 55-85.

Belkin, N.J. Oddy, R.N. & Brooks, H.M. (1982a) ASK for information retrieval: part I. **Journal of Documentation**, 33(2), pp. 61-71.

Belkin, N.J. Oddy, R.N. & Brooks, H.M. (1982b) ASK for information retrieval: part II. Results of a design study, **Journal of Documentation**, 38(3), pp. 145-164.

Bush, V. (1945) As we may think. **The Atlantic Monthly**, 176(1), pp. 101-108.

Byström, K. & Järvelin, K. (1995) Task complexity affects information seeking and use. **Information Processing & Management**, 31(2), pp. 191-213.

Card, S. Moran, T. & Newell, A. (1983) **The psychology of human-computer interaction**. Hillsdale, NJ: Erlbaum.

Cowie, J. & Lehnert, W. (1996) Information extraction. **Communications of the ACM**, 39(1), pp. 80—91.

Darke, J. (1978) The primary generator and the design process. In: Rogers, W.E. & Ittelson, W.H. eds. **New directions in environmental design research: proceedings of EDRA 9**. Washington: EDRA. pp. 325-337.

De Mey, M. (1980) The relevance of the cognitive paradigm for Information Science. IN: Harbo, O. ed. **Theory and application of information research**, London, Mansell. pp. 48-61.

Dervin, B. (1983) An overview of sense-Making research: Concepts, methods and results. **Paper presented at the Annual Meeting of the International Communication Association, Dallas, TX, May 1983**. [online] Available:

http://communication.sbs.ohio-state.edu/sense-making/art/artabsdervin83smover view.html [Accessed October 28, 2003]

Dervin, B. (1994) Information <-> democracy: An examination of underlying assumptions. **Journal of the American Society for Information Science**, 45(6), pp. 369-385.

Dervin, B. Mahey, P.J. Osborne, T. Huesca, R. & Higgins, J. (1992) The problematics of dialogue: Some ideas in the process of becoming. **Paper presented at the International Association of Mass Communication Research conference, Sao Paulo, Brazil, August 1992.**

Dervin, B. & Nilan, M. (1986) Information needs and uses. **Annual Review of Information Science and Technology**, 21, pp. 3-33.

Dey, I. (1999) **Grounding grounded theory: Guidelines for qualitative enquiry**. San Diego, Academic Press.

Donaldson, S. (1999) Predictive learning and cognitive momentum: A foundation for intelligent, autonomous systems. IN: **Proceedings of the 37th Annual ACM Southeast Regional Conference, Mobile, AL, USA, April 1999**, ACM.

Drori, O. (2003) How to display search results in digital libraries–user study. IN: **Proceedings of the 3rd International Workshop on New Development in Digital Libraries, NDDL 2003, Angers, France, April 2003**, Portugal: ICEIS Press, pp. 13-28.

Ellis, D. (1989a) A behavioural approach to information retrieval system design. **Journal of Documentation**, 45(3), pp. 171-212.

Ellis, D. (1989b) A behavioural model for information retrieval system design. **Journal of Information Science**, 15, (4/5), pp. 237-247.

Ellis, D. (1990) **New horizons in information retrieval**. London, Library Association.

Ellis, D. (1992) The physical and cognitive paradigms in information retrieval research. **Journal of Documentation**, 48(1), pp. 45-64.

Ellis, D. Cox, D. & Hall, K.A. (1993) Comparison of the information seeking patterns of researchers in the physical and social sciences. **Journal of Documentation**, 49(4), pp. 356-369.

Ellis, D. & Haugan, M. (1997) Modelling the information seeking patterns of engineers and research scientists in an industrial environment. **Journal of Documentation**, 53(4), pp. 384-403.

Eastman, C.M. (1970) On the analysis of the intuitive design process. IN: Moore G.T. ed. **Emerging methods in environmental design and planning**, Cambridge, Mass, MIT Press. pp. 21-37.

Farradane, J. (1970) The institute: The first twelve years. **Information Scientist**, 4, pp. 143-151.

Fuhr, N. (1992) Probabilistic Models in Information Retrieval. **The Computer Journal**, 35 (3), pp. 243-255.

Frohmann (1992) The power of images: A discourse analysis of the cognitive viewpoint. **Journal of Documentation**, 48(4), pp. 365-386.

Gaizauskas, R. & Wilks, Y. (1998) Information extraction: Beyond document retrieval. **Journal of Documentation**, 54(1), pp. 70-105.

Glaser, B. & Strauss, A. (1967) **The discovery of grounded theory: Strategies for qualitative research**. Chicago, Aldine.

Goel, V. & Pirolli, P. (1992) The structure of design problem spaces. **Cognitive Science**, 16, pp. 395-429.

Haake, M. & Wilson, B. (1992) Supporting Collaborative Writing of Hyperdocuments in SEPIA. **Proceedings of ACM CSCW'92 Conference on Computer-Supported Cooperative Work**, pp.138-146.

Haas, C (1989) Does the medium make a difference? Two studies of writing with pen and paper and with computers. **Human-Computer Interaction** 4(2) pp.149-169.

Halasz, F.G. Moran, T.P. & Trigg, R.H. (1987) NoteCards in a nutshell, **Proceedings of ACM CHI+GI'87 Conference on Human Factors in Computing Systems and Graphics Interface**, pp. 45-52.

Hansen, K.A. & Ward, J. (1991) Information Technology: Changes in Large Newspaper Libraries, **Special Libraries** (Fall), pp. 267-273.

Hayes, L.S. & Flower, J.R. (1980) Identifying the organization of writing processes. IN: Gregg, L.W. & Steinberg, E.R. eds. **Cognitive Processes in Writing**, Hillsdale, NJ, Lawrence Erlbaum, pp. 3-30.

Hendry, D.G. & Harper, D.J. (1997) An informal information-seeking environment. **Journal of the American Society for Information Science**, 48(11), pp. 1036-1048.

Hewett, T.T. Baecker, R. Card, S. Carey, T. Gasen, J. Mantei, M. Perlman, G. Strong, G. & Verplank, W. (2002) **ACM SIGCHI Curricula for Human-Computer Interaction** [online] Available: http://sigchi.org/cdg/index.html [Accessed October 24, 2003]

Hollan, J. Hutchins, E. & Kirsh, D. (2000) Distributed Cognition: Towards a new foundation for human-computer interaction research. **ACM Transactions on Computer-Human Interaction**, 7(2), pp. 174-196.

Ingwersen, P. (1982) Search procedures in the library—analysed from the cognitive point of view. **Journal of Documentation**, 38(3), pp. 165-191.

Ingwersen, P. (1992) **Information Retrieval Interaction**. London, Taylor Graham. [online] Available: http://www.db.dk/pi/iri/ [Accessed October 28, 2003]

Ingwersen, P. (1996) Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory. **Journal of Documentation**, 52(1), pp. 3-50.

Ingwersen, P. & Pors, N.O. eds. (1995) Information and Information Science. IN: **Encyclopedia of Library and Information Science**, 56, Suppl. 19, New York, Marcel Dekker Inc. pp. 137-174.

Isokoski, P. (2001) Model for unistroke writing time: the write stuff. **Proceedings of ACM CHI 2001 Conference on Human Factors in Computing Systems**, pp.357-364.

Kellogg, R.T. & Mueller, S. (1993) Performance amplification and process restructuring in computer-based writing. **International Journal of Man-Machine Studies**, 39(1), pp.33-49.

Kotonya, G. & Sommerville, I. (1998) **Requirements Engineering: Processes and Techniques**. John Wiley & Sons, Inc. New York, NY, USA.

Kuhlthau, C.C. (1988a) Developing a model of the library research process: investigations of cognitive and affective aspects. **Reference Quarterly**, 28, pp. 232-242.

Kuhlthau, C.C. (1988b) Perceptions of the information search process of users in libraries: a study of changes from high school through college. **Information Processing and Management**, 24, pp. 419-427.

Kuhlthau, C.C. (1988c) Longitudinal case studies of the information search process of users in libraries. **Library and Information Science Research**, 10, pp. 251-304.

Kuhlthau, C.C. (1989) The information search process of high-middle-low achieving high school seniors. **School Library Media Quarterly**, 17, pp. 251-304.

Kuhlthau, C.C. (1991) Inside the search process: Information seeking from the user's perspective. **Journal of the American Society for Information Science**, 52(5), pp. 361-371.

Kuhlthau, C.C. (1993) A principle of uncertainty for information seeking. **Journal of Documentation**, 49(4), p.39-55.

Kuhlthau, C.C. (1999) Accommodating the user's information search process: Challenges for information retrieval system designers. **Bulletin of the American Society for Information Science**, 25(3), pp. 12-16.

Kuhlthau, C.C. & Tama, S.L. (2001) Information Search Process of Lawyers: A Call for 'Just for me' information services. **Journal of Documentation**, 57(1), pp. 25-43.

Kuhlthau, C.C. Turock, B. George, M. & Belvin, R. (1990) Validating a model of the search process: a comparison of academic, public and school library users. **Library and Information Science Research**, 12, pp. 5-32.

Lawson, B. (1997) **How designers think: The design process demystified**, 3rd ed. Oxford, Butterworth-Heinemann.

Line, M. (1969) Information requirements in the social sciences: some considerations. **Journal of Librarianship**, 1(1), pp. 1-19.

Line, M. (1974) Draft definitions: information and library needs, wants, demands and uses. **Aslib Proceedings**, 26(2), p. 87

Lueg, C. (2002) On problem solving and information seeking. IN: **Proceedings of Information Seeking in Context: The Fourth International Conference on**

Information Needs, Seeking and Use in Different Contexts, 11-13 September 2002, Lisbon, Portugal. [online] Available: http://www-staff.it.uts.edu.au/~lueg/papers/isic02s.pdf [Accessed October 28, 2003]

Luhn, H. (1960) Keyword in context index for technical literature, **American Documentation**, 9(4), p.288-295.

Marchionini, G. (1995) **Information seeking in electronic environments**. Cambridge, UK, Cambridge University Press.

Martin, H. & Nicholas, D. (1993) End-users coming of age? Six years of end-user searching at The Guardian. **Online & CDROM Review**, 17(2).

McFarlane, D.C. (1998) **Interruption of people in human-computer interaction**. PhD Doctoral Dissertation, George Washington University, Washington, DC.
Mitchell, A. Posner, I. & Baecker, R. (1995) Learning to Write Together Using Groupware Papers: **Proceedings of ACM CHI'95 Conference on Human Factors in Computing Systems**, pp. 288-295.

Miura, T. (1996) **Behavior and Visual Attention**. Japan, Kazama.

Neuwirth, C.M. Chandhok, R. Charney, D. Wojahn, P. & Kim, L. (1994) Distributed collaborative writing: A comparison of spoken and written modalities for reviewing and revising documents. **Proceedings of ACM CHI'94 Conference on Human Factors in Computing Systems**, 1, pp. 51-57.

Newell, A. (1981) The knowledge level, **The AI Magazine**, 2(2), pp. 1-20.

Ng, K.B. (2002) Toward a theoretical framework for understanding the relationship between situated action and planned action models of behaviour in information retrieval contexts: contributions from phenomenology. **Information Processing and Management**, 38, pp. 613-626.

Nicholas, D. (1996a) An assessment of the online searching behaviour of practitioner end users, **Journal of Documentation**, 52(3), pp. 227-251.

Nicholas, D. (1996b) **Assessing information needs: tools and techniques**, London, Aslib.

Nicholas, D. & Martin, H. (1997) Assessing information needs: a case study of journalists. **Aslib Proceedings**, 49(2), pp. 43-52.

Nicholas, D. Williams, P. Cole, P. & Martin, H. (2000) The impact of the Internet on information seeking in the Media. **Aslib Proceedings**, 52(3). pp. 98-114.

Offer, D. (2003) How to display search results in digital libraries - user study, **Proceedings of the 3rd International Workshop on New Development in Digital Libraries**, *NDDL 2003,* Angers, France, April 2003, ICEIS Press, pp. 13-28.

O'Malley, K. & Sharples, M. (1986) Tools for management and support of multiple contstraints in a writer's assistant, **Proceedings of the HCI'86 Conference on People and Computers III.** pp.115-131.

Paisley, W. (1968) Information needs and uses. **Annual Review of Information Science and Technology**, 3, pp. 1-30.

Payne, S. & Green, T. (1986) Task-Actions Grammars: A Model of the Mental Representation of Task Languages. **Human-Computer Interaction**, 2, pp.93-133.

Rasmussen, J. Pejtersen, A.M. & Goodstein, L.P. (1994) **Cognitive Systems Engineering**. New York, Wiley.

Reitman, W.R. (1964) Heuristic decision procedures, open constraints, and the structure of ill-defined problems. IN: Shelley, M. & Bryan, G. eds. **Human Judgements and Optimality**, New York, Wiley.

Reitz, J.M. (2002) **ODLIS: Online Dictionary of Library and Information Science** [Online] Available: http://www.wcsu.edu/library/odlis.html [Accessed October 24, 2003]

Robertson, S.E. & Sparck-Jones, K. (1976). Relevance weighting of search terms. **Journal for the American Society of Information Science**, 27(3), pp.129-146.

Rowe, P.G. (1987) **Design thinking**. Cambridge, Mass, MIT Press.

Saal, F.E. Downey, R.G. & Lahey, M.A. (1980) Rating the ratings: Assessing the psychometric quality of rating data. **Psychological Bulletin**, 88(2). pp.413-428.

Salton, G. (1968) **Automatic information organisation and retrieval**. New York McGraw-Hill.

Sanna, T. (1997) Constituting 'information' and 'user' as research objects: A theory of knowledge formations as an alternative to the information man-theory. IN: Vakkari, P. Savolainen, R. & Dervin, B. eds. **Information seeking in context (ISIC): Proceedings of an international conference on research in information needs, seeking and use in different contexts**, 14-16 August, 1996, Tampere, Finland, London, Taylor Graham. pp. 67-80.

Scardamalia, M. & Bereiter, C. (1987) Knowledge telling and knowledge transforming in written composition. IN: Rosenberg, S. ed. **Advances in applied psycholinguistics**, Cambridge: Cambridge University Press. pp. 143-175.

Schön, D.A. (1983) **The reflective practitioner: How professionals think in action**. New York, Basic Books.

Schuler, W. & Smith, J.B. (1990) Author's Argumentation Assistant (AAA): A hypertext-based authoring tool for argumentative texts. **Proceedings of the ECHT'90 European Conference on Hypertext**, pp.137-151.

Searle, J. (1984) **Minds, brains and science**. Cambridge, MA, Harvard University Press.

Semonche, B.P. (1993) News Library History, IN: Semonche, B.P. ed. **New media libraries: a management handbook**, (Greenwood Library Management Collection), Westport, CT: Greenwood Press.

Senn, S. (1997) **Statistical issues in drug development**, Chichester, UK, Wiley.

Sharples, M. (1996) An account of writing as creative design. IN: Levy, C.M., & Ransdell, S. eds. **The science of writing**. New Jersey, Lawrence Erlbaum.

Simon, H.A. (1973) The structure of ill-structured problems. **Artificial Intelligence**, 4, pp. 181-204.

Simpson, R. Renear, A. Mylonas, E. & van Dam, A., (1996) 50 years after 'As We May Think: The Brown/MIT Vannevar Bush Symposium, **ACM Interactions**, 3(2).

Spink, A. Wilson, T.D. Ford, N. Foster, A. & Ellis, D. (2002) Information seeking and mediated searching study, Part 3, successive searching. **Journal of The American Society for Information Science and Technology**, 53(9), pp. 716-727.

Stephens, M. (1993) **Broadcast news**. (3rd ed.), Fort Worth, Harcourt Brace Jovanovich.

Strauss, A. & Corbin, J. (1990) **Basics of qualitative research: Grounded theory procedures and techniques**. 1st ed. London, Sage.

Strauss, A. & Corbin, J. (1998) **Basics of qualitative research: Techniques and procedures for developing grounded theory**. 2nd ed. London, Sage.

Suchman, L. (1987) **Plans and situated actions: The problem of human machine communication**. Cambridge, UK, Cambridge University Press.

Sullivan, N. (1993) Ask not for whom the phone rings: It rings for thee. **Home-office computing**, 11(3), pp. 104

Talja, S. (1997) Constituting 'information' and 'user' as research objects: a theory of knowledge formations as an alternative to the information-man theory. IN: Vakkari, P. Savolainen, R. & Dervin, B. eds. **Information seeking in context (ISIC): Proceedings of an international conference on research in information needs, seeking and use in different contexts**, 14-16 August, 1996, Tampere, Finland, London, Taylor Graham. pp. 67-80.

Tang, R. & Solomon, P. (1998) Toward an understanding of the dynamics of relevance judgment: An analysis of one person's search behaviour, **Information Processing and Management**, 34(2-3), pp. 237-256.

Taylor, R.S. (1968) Question-negotiation and information seeking in libraries. **College and Research Libraries**, 29, 178-194.

Thomas, C. (1987) Designing electronic paper to fit user requirements. **Proceedings of the HCI'87 conference on people and computers III**. pp.247-257.

Tiamiyu, M.A. (1992) The relationships between source use and work complexity, decision-maker discretion and activity duration in Nigerian government ministries. **International Journal of Information Management**, 12, pp. 130-141.

Tuominen, K. & Savolainen, R. (1997) Social constructionist approach to the study of information use as discursive action. IN: Vakkari, P. Savolainen, R. & Dervin, B. eds. **Information seeking in context (ISIC): Proceedings of an international conference on research in information needs, seeking and use in different contexts**, 14-16 August, 1996, Tampere, Finland, London, Taylor Graham. pp. 81-96.

Vakkari, P. (1998) Growth of theories on information seeking: An analysis of growth of a theoretical research program on the relation between task complexity and information seeking. **Information Processing & Management**, 34(3-4), pp. 361-382.

Vakkari, P. (1999) Task complexity, problem structure and information actions: Integrating studies on information seeking and retrieval. **Information Processing and Management**, 35, pp. 819-837.

Vakkari, P. (2000a) Cognition and changes of search terms and tactics during task performance: a longitudinal study. IN: **Proceedings of the RIAO 2000 conference**, Paris, CID. pp. 894-907.

Vakkari, P. (2000b) Relevance and contributory information types of searched documents in task performance. IN: **Proceedings of SIGIR 2000 Conference**, Athens, New York, ACM, pp. 2-9.

Vakkari, P. (2001) A theory of the task-based information retrieval process: a summary and generalisation of a longitudinal study. **Journal of Documentation**, 57(1), pp. 44-60.

Vakkari, P. & Hakala, N. (2000) Changes in relevance criteria and problem stages in task performance. **Journal of Documentation**, 56(5), pp. 540-562.

Vakkari, P. & Pennanen, M. (2001) Sources, relevance and contributory information of documents in writing research proposal: a longitudinal case study. IN: Höglund, L. & Wilson, T.D. eds. **The new review of information behaviour research: studies of information seeking in context (Proceedings of ISIC 2000)**, Cambridge, Taylor Graham. pp. 217-232.

Van de Ven, A. & Ferry, D. (1980) **Measuring and assessing organizations**. New York, Wiley.

Walker, J.H. (1988) The role of modularity in document authoring systems, **Proceedings of the ACM conference on document processing systems**, p.117-124.

Weizenbaum, J. (1966) ELIZA: a computer program for the study of natural language communication between man and machine. **Communications of the ACM**, 9(1), pp. 36-35.

Wersig, G. (1979) The problematic situation as a basic concept of Information Science in the framework of social sciences: A reply to Belkin, N.J. IN: **Inter-national Federation of Documentation. New trends in informatics and its terminology**. Moscow, VINITI, 1979. pp. 48-57.

Wilson, T. (1996) **Information behaviour: an interdisciplinary perspective**, A report to the British Library Research & Innovation Centre on a review of the literature. [Online] Available: http://informationr.net/tdw/publ/infbehav/prelims.html.

Wilson, T. (1999) Models of information behaviour research. **Journal of Documentation**, 55(3), pp. 249-270.

Wilson, T. (2000) Human information behavior. **Informing Science**, 3(2) (special issue on information science research), pp. 49-55.

Wilson, T.D. Ford, N. Ellis, D. Foster, A. & Spink, A. (2002) Information seeking and mediated searching. Part 2. Uncertainty and its correlates. **Journal of the American Society for Information Science and Technology**, 53(9), pp. 704-715.

Woods, D.D. Johannesen, L. Cook, R.I. & Sarter, N.B. (1994) **Behind human error: Cognitive systems, computers and hindsight**. Dayton, OH: CSERIAC*Jiajie Zhang* (The Ohio State University), *Donald A. Norman* (Apple Computer, Inc.), *Cognitive Science* ,18, pp. 87-122.

Yang, S.C. (1997) Information seeking as problem-solving using a qualitative approach to uncover the novice learner's information–seeking processes in a Perseus hypertext system. **Library & Information Science Research**, 19(1), pp. 71-92.

Appendices

Appendix I

Exploratory study - Subject Instructions

Its 1.30 in the afternoon. You are a journalist working in the newsroom of a fast moving national newspaper. The newswire, that provides a very brief outline of breaking events, reports the following

Footballer Jeff Mackenzie spent last night in the cells after being arrested for assaulting actress Kate Whitely at Kix nightclub. Whitely was taken to casualty but was released soon after. It is not known if she is pressing charges.

The editor of the paper (who is infamous for his rather 'direct' approach to staff management) has sent a reporter to gather more information, but he has told you that he wants you to write a piece (about 300-400 words) based on what is known already about Mackenzie and Whitely. He tells you that Mackenzie has quite a reputation for bad behaviour.

The only resource for research that you have is the world-wide-web. You have a search intermediary (the experimenter) to help you find resources and perform searches. The resources that he is familiar with are:-

Financial Times Global Archive	Search an archive of world newspapers, trade publications, magazines and newsletters, wire services and others	1996 to current
Guardian Unlimited	Search the Guardian and Observer archive.	Sept 1998 to current
The Times and Sunday Times	Browse an archive of internet editions. Select by date only.	
ITN news archive	Search the archive of ITN news stories	Feb 99 to current
BBC News archive	Search the archive of BBC news stories (advanced search)	Nov 1997 to current
Hansard	Search the archive of transcripts of House of Commons debates	Nov 93 to current
Google	Search the worldwide web.	current
Altavista	Search the worldwide web (advanced search)	current
Excite-UK	Search UK or European web sites	current

During your research you will be able to see the search intermediary's monitor so that you can tell him which resource you would like to use and what documents/information you would like to find (and even how to find it). The first thing that this monitor will show is the table of resources with links as shown above. When

you see a document that you want to read you can either read it on the screen or have it printed.

You will also see another monitor in front of you (with a keyboard) running Microsoft Word. This is where you should write your report and any notes. You will find two files open – one for notes, and one for your report.

One last thing - before marching back to his office, the editor barks that he wants the report finished in two hours at the most.

Appendix II

Appendix IIa[1]

(Report)
Footballer Jeff Mackenzie went off the rails yet again last night, as it was reported that he assaulted Rescue actress Kate Whitely at the famous Soho nightspot Kix. Mackenzie , 29, spent the night in police cells, and now waits to hear if Kate, 24, will press charges against him for his latest misdemeanour. The attack comes as the latest in a long line of unsavory incidents involving Mackenzie and mirrors his previous attack on then girlfriend, TV celebrity Justine Carpenter, in a crowded Madrid bar during the World Cup finals in the summer of 1994. Kate, like Justine, is one of television's hottest properties. She stars in the hit BBC drama Rescue and has been previously voted one of the top 100 'sexiest women in the world' by 'lad's mag' LM.

News of the assault will come as no surprise to either Justine or former girlfriend Carol While, who had Mackenzie charged with assaulting her during Christmas 1993. Mackenzie was acquitted of those charges, but he may not escape punishment for his latest show of brutality. It will also come as no surprise to anyone who has been following Mackenzie 's dramatic fall from grace.

Mackenzie , who has recently shown signs of recovering from the illness that he claims almost caused him to take his own life during his time at Rovers is almost certain to be given his marching orders by new boss, Wanderer's David Roane. Mackenzie was already on his last chance at Wanderers, following his notorious 'champagne bottle' escapade in Barcelona and reports of violence in a bar just hours before that game.

The football career that shot Jeff Mackenzie into the public eye during the mid-nineties now lies in tatters. He was one of football's hottest properties when City signed him from United for a British transfer record back in 1993, but had already acquired a 'bad boy' tag, following rumours of dressing room unrest. Despite a moderately successful spell with City, his inability to live with the pressure of being one of the game's top players lead to mounting disciplinary problems and a £7million transfer to Rovers, in his native Midlands. It was while at Rovers that his problems began to spiral out of control, ending with a stay in Guildford's famous Chapel clinic, where he was treated for severe depression. Mackenzie often commented on the lack of understanding afforded his condition from fellow professionals and supporters alike, but was accused of taking cocaine and marijuana while cavorting with fellow female patients during his time there. It was during these days that he contemplated suicide. Less than a month ago Rovers manager David Beecham offloaded Mackenzie for the cut-price fee of £250,000 to Wanderers, where he hoped to rebuild his career.

[1] Some dates, names and roles have been changed in order to protect the patently innocent.

Appendix IIb[2]

(report)
Jeff Stikes Another Stunna

Disgraced footballer Jeff Mackenzie spent last night in the cells after being arrested for assaulting an actress Kate Whitely at a nightclub. Whitely was immediately rushed to casualty. Kate who plays TV doctor Karin West in the drama series Rescue was on an evening out with friends at exclusive Kix's nightclub in London's West End when the assault allegedly took place.

Mackenzie is quoted as saying "Maybe in this circumstance, I'll hold my hands up and say 'Naive' but at the end of the day, I haven't killed anybody".

Sexy Stunna Kate who shot to fame with her nude glamour modelling, is alleged to have recently split with boyfriend Paul Morrison who also stars in the series as gay nurse Pete. announced that she is quitting the hit TV series. Rumour has it she recently started a relationship with boozy footballer Jeff.

Jeff Mackenzie is no stranger to controversy. After reeking havoc in a German Bar last month which resulted in the teams deportation, Mackenzie is currently completing 'community service' imposed by Wanderers' boss David Roane at Birmingham school

He has also been linked to several attacks on women which he has admitted were fuelled by drink. Former news reader and Head to Head star Justine Carpenter was on the recieving end of crazed Mackenzie 's wrath in 1990 during the World Cup Tournament in Madrid, Jeff caught her drinking with his footballing pal Craig Norman, and flew into a rage. When she refused to leave with him he allegedly dragged her to the floor and aimed kicks at her head. Staff promptly threw him out, as did Justine shortly after, Although charged with assault the case against Mackenzie was dropped through lack of evidence. But even this was not the first time Jeff has been linked with battering a girlfriend. In 1998, he was cleared of assaulting a former girlfriend, Carol While during a late night visit.

Mackenzie 's career has taken a downhill path since making his record breaking move to City in 1995. In 1997 he moved to Rovers which proved to be the lowest point in his career until now, including culminating in his one-goal disaster loan at first division Albion. His poor performance on the pitch led to treatment at the world famous Chapel Clinic for depression, a condition often linked to excessive drinking. Football pundit Rob Smith branded him as 'not fit to wear the famous green and blue colours of Rovers'.

Whitely was released soon after being admitted to casualty. It is not known if she is presing charges.

[2] Some dates, names and roles have been changed in order to protect the patently innocent.

Appendix III

Appendix IIIa

Questionnaire initially used for second round field study interviews

Seeking information for originality checking with respect to published articles

a) Do you check that no one else has published a piece you are intending to publish before you start work on it?

If yes...

b) What aspect of an idea is it that you check for originality?
c) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
d) Why might you do this? How does it help you?
e) How do you do this?

Ancillary questions

f) Is it important?

If no...

g) How much would you like to be able to do this?
h) Under what specific circumstances do you think this would be useful to you?

Seeking information for originality checking with respect to current assignments in the organisation

a) Do you check that no one else is producing a piece at the Times that overlaps with yours?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions

e) Do you do this often?

If no...

f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Seeking an overview of the background of an issue

a) Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions

e) Do you do this often?

If no...

f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Seeking an overview of the background on a person

a) Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions

e) Do you do this often?

If no...

f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Seeking an overview on the background to a scientific issue

a) Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions

e) Do you do this often?

If no...

f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Seeking profiles of companies

a) Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions

e) Do you do this often?

If no...

f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Seeking profiles of people

a) Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions

e) Do you do this often?

If no...

f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Finding book reviews

a) Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions

e) Do you do this often?

If no...

f) How much would you like to be able to do this?

g) Under what specific circumstances do you think this would be useful to you?

Seeking facts and figures to support a chosen angle

a) Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions

e) Do you do this often?

If no...

f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Seeking the causes of a past disasters or catastrophes of a given type

a) Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions

e) Do you do this often?

If no...

f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Finding out where past disasters or catastrophes of a given type happened

a) Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions
e) Do you do this often?

If no...
f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Finding the numbers of dead and injured in past disasters or catastrophes of a given type

a) Do you ever do this?

If yes...
b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions
e) Do you do this often?

If no...
f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Finding out if someone has previous convictions

a) Do you ever do this?

If yes...
b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions
e) Do you do this often?

If no...
f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Finding out book titles by specific people

a) Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions

e) Do you do this often?

If no...

f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Finding historical quotations

a) Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions

e) Do you do this often?

If no...

f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Finding non-historical quotations

a) Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions

e) Do you do this often?

If no...

f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Confirming quotation wording

a) Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions

e) Do you do this often?

If no...

f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Seeking lines of dialogue from films

a) Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions

e) Do you do this often?

If no...

f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Establishing the credibility of un-sourced information

a) Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions

e) Do you do this often?

If no...

f) How much would you like to be able to do this?

g) Under what specific circumstances do you think this would be useful to you?

Checking the authority of a web site

a) Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions

e) Do you do this often?

If no...

f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Checking names and their spellings

a) Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions

e) Do you do this often?

If no...

f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Checking facts with article writer

a) Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions

e)Do you do this often?

If no...

f) How much would you like to be able to do this?

g)Under what specific circumstances do you think this would be useful to you?

Seeking useful contacts/informants

a)Do you ever do this?

If yes...

b)Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?

c)Why might you do this? How does it help you?

d)How do you do this?

Ancillary questions

e)Do you do this often?

If no...

f) How much would you like to be able to do this?

g)Under what specific circumstances do you think this would be useful to you?

Finding journalists who have written on a particular subject in the past

a)Do you ever do this?

If yes...

b)Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?

c)Why might you do this? How does it help you?

d)How do you do this?

Ancillary questions

e)Do you do this often?

If no...

f) How much would you like to be able to do this?

g)Under what specific circumstances do you think this would be useful to you?

Finding the name of an 'expert' in a particular field

a)Do you ever do this?

If yes...

b)Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?

c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions
e) Do you do this often?

If no...
f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Finding organizations to contact

a) Do you ever do this?

If yes...
b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions
e) Do you do this often?

If no...
f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Seeking contact details for a company

a) Do you ever do this?

If yes...
b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions
e) Do you do this often?

If no...
f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Finding contact detail for an individual

a) Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions

e) Do you do this often?

If no...

f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Thorough/exhaustive information seeking

a) Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions

e) Do you do this often?

If no...

f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Making online notes

a) Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions

e) Do you do this often?

If no...

f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Making offline notes

a) Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions

e) Do you do this often?

If no...

f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Transferring text from cuttings to 'basket'

a) Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions

e) Do you do this often?

If no...

f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Transferring text from cuttings directly into copy

a) Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions

e) Do you do this often?

If no...

f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Printing cuttings

a) Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions

e) Do you do this often?

If no...

f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Printing web pages

a) Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions

e) Do you do this often?

If no...

f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Highlighting text in printed cutting

a) Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions
e)Do you do this often?

If no...
f) How much would you like to be able to do this?
g)Under what specific circumstances do you think this would be useful to you?

Storing information for the medium to long term

a)Do you ever do this?

If yes...
b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c)Why might you do this? How does it help you?
d)How do you do this?

Ancillary questions
e)Do you do this often?

If no...
f) How much would you like to be able to do this?
g)Under what specific circumstances do you think this would be useful to you?

Relocating local electronic documents

a)Do you ever do this?

If yes...
b)Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c)Why might you do this? How does it help you?
d)How do you do this?

Ancillary questions
e)Do you do this often?

If no...
f) How much would you like to be able to do this?
g)Under what specific circumstances do you think this would be useful to you?

Relocating information in online cuttings

a)Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions
e) Do you do this often?

If no...
f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Relocating information in printed cuttings

a) Do you ever do this?

If yes...
b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions
e) Do you do this often?

If no...
f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Writing a short structure plan

a) Do you ever do this?

If yes...
b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions
e) Do you do this often?

If no...
f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions
e) Do you do this often?

If no...
f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Relocating information in printed cuttings

a) Do you ever do this?

If yes...
b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions
e) Do you do this often?

If no...
f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Writing a short structure plan

a) Do you ever do this?

If yes...
b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions
e) Do you do this often?

If no...
f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Using a structure plan as a content checklist

a) Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions

e) Do you do this often?

If no...

f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Using highlighted cuttings as a content checklist

a) Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions

e) Do you do this often?

If no...

f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Using chronology as a content checklist

a) Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions

e) Do you do this often?

If no...

f) How much would you like to be able to do this?

g) Under what specific circumstances do you think this would be useful to you?

Placing printouts in easily accessible location

a) Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions

e) Do you do this often?

If no...

f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Placing 'basket' in easily accessible location

a) Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions

e) Do you do this often?

If no...

f) How much would you like to be able to do this?
g) Under what specific circumstances do you think this would be useful to you?

Placing a structure plan in an easily accessible location

a) Do you ever do this?

If yes...

b) Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?
c) Why might you do this? How does it help you?
d) How do you do this?

Ancillary questions

e)Do you do this often?

If no...

f) How much would you like to be able to do this?

g)Under what specific circumstances do you think this would be useful to you?

Changing plans

a)Do you ever do this?

If yes...

b)Under what specific circumstances do you do this? (if many) Under what circumstances wouldn't you do this?

c)Why might you do this? How does it help you?

d)How do you do this?

Ancillary questions

e)Do you do this often?

If no...

f) How much would you like to be able to do this?

g)Under what specific circumstances do you think this would be useful to you?

Appendix IIIb

Field study interview transcript - Lewis Smith

Res I notice that you work on breaking news items most of the time, so perhaps it would be useful to talk about that.

LS Yeah, I mean I just put 1% on features which is [unint] on the that [unint] the Times do set [unint]

Res Maybe new stories are more interesting because they are more time critical in terms of finding information.

LS They are for me.

Res What I wanted to do is get you to outline the process of writing a breaking news item.

LS The news desk tells you what they want. Sometimes you have to work with somebody else; usually you are on your own. They tell you the story, sometimes they will tell you the line that they are particularly interested in. And you then gather information, write the story and send it back to them.

Res What sort of information might you gather?

LS Obviously it depends on the story.

Res Is there a particular story that you have done recently?

LS Nothing that springs to mind at the moment. Say it was a train crash for example, the first thing you do is to find out if anybody had died and/or injury, and how serious it was from the point of view of the reaction of the emergency services. And you would also get to take into account the wider implications of transport, whether this was their second train crash in two weeks or the fifth in two months, and whether there is a common cause, whether they were all...

Res So it is identifying patterns?

LS Widely, yes. But what you really want to be able to do is tell the reader exactly why this crash happened and what can be done about it to stop it happening again, and then you will read the accident investigator's report. The main aim is to tell the reader what is happening. So first of all you say '2 people died, 7 people injured when 2 trains collided'. And the next thing you have to say is how this affects people there, the immediate effects, and then you have got to take the long-term, whether the rail firms have to, built-in older [unint] replacing, their safety procedures. Really everything

you can, and then simply [unint]. A big story like that absolutely got to get eyewitness reports. Preferably survivors and after that you start thinking about witnesses. I know that the [unint] saw the crash, if you are luckily enough to find them, and the next slot is for the emergency workers on the scene, because that always creates a human element which is [unint].

Res With a story like that, what sort of information might you look for in past news wires and news reports.

LS From archive material?

Res Archive material.

LS You would want first of all to see if there were links on cause or a common geography. And then you start looking at things like, well, let's give the reader a box of all the train crashes in the last ten years, or 'is this the biggest crash ever?' which is like record breaking features about it, like was it the most deaths in Britain, the most injured, anything like that. That is why you go to archive. You would also use them for reference. It never harms, certainly the general report, to be reminded how in the past these stories have been treated. It reminds you of an angle... you have got to think of a good thing, and bearing in mind, say if a crash happens a lunchtime, you have got to have copy ready five minutes later. And although you can think of a lot of obvious things, there might be other suggestions in archive material if you get the chance to read them, that you think 'that is a good idea, let's see if we can get anywhere with that, that line of question'. That would probably be more useful for day 2, but that is certainly, I have found...

Res With day 2 being a follow-up longer piece.

LS Well no, not necessarily longer, but more detailed. I mean if we are talking about the train crash itself, certainly on the first day, you would want to get as much eyewitness report as possible. You just want to hit the reader with the horror of what happened, the human horror, and day 2, you usually use a witness... you will get witnesses, you might get the first people if possible, before they are able to talk. You might get some particularly interesting witnesses fall out of the woodwork. Or there might be somebody who wakes up, who is on death's door, wasn't expected to survive and there is always lines like that that can be approached, and that might happen a few days later. But the way things are going, you start... start looking at the rail industry, serious questions, 'Why did this happen?', 'What have the rail companies done about it?'... more considered pieces, where you have got over the initial rush, the initial adrenalin. The first day you tend to know what has happened anyway. They know two trains have crashed. By the end of the day

you might have a line that investigators suspect it was a red signal that the train had gone through, or that it was a cracked piece of track. I think that is accurate, on the first day I think we had a pretty good idea that it was a fact that it was the track. And the second day, that always becomes a damn sight clearer, where the investigators are in a position where they can rule out certain factors. They are also able to go into more detail about what they know, what they have picked up on. They can then reassess it themselves, and the transport correspondent on a case like that has then got to use specialised contacts to come up with more detail... to have expert opinions.

Res If you were the journalist who was assigned to write the breaking story, would you then as a matter of course be the person who wrote the follow-up.

LS No, not necessarily. I mean, you would usually expect to go back on the story the next day, but it doesn't necessarily happen. It depends what reporters they have got available, what your diary is looking like and someone might be off. The news-desk will say 'it's fine, we have got enough people'. If you got sent out at the scene, certainly chances are that you would stay overnight and be back there the next morning.

Res When you are collecting the archive material, where would you go to do that?

LS I tend to use a mixture of the editorial database and Lexus Nexus. Sometimes paper cuts which are often extremely useful but don't give [unint].

Res These are hard copies.

LS Yeah, its quite literally cuttings of newspaper reports. I find these useful from time to time. Archives are pretty reasonable, but I don't have to trust to look, but I [unint]. You sometimes find it difficult to narrow it down. There are very stupid little things, like I have used it for all different things because when I use Lexus Nexus I can't actually remember if I put the word 'and' between words, just a space or a semicolon. I either do it by trial and error or ask the library to do it. And that is a bigger problem with the editorial database, because I learnt that system a couple of years ago. Somebody tried to teach me the other way of [unint]. Nobody has actually told me why.

Res The other way being?

LS Well you put commands in, like 'g' for get. Somebody gave me a worksheet but it didn't make sense, certainly not when you are in a hurry and you want a quick... it is not ABC. And so I scrapped that

and do it as I used to do, but the way I used to do it doesn't seem to have any easy facility for bracketing the date. I am sure somebody has told me this at some point, how I can do a date, but I can't remember, and I don't have a sort of easy reference sheet to go and look at it.

Res Adding a date parameter into the query is quite useful?

LS Well if you want all train crashes, you put train crashes. There are hundreds and hundreds of items that the system either can't cope with or you can't cope with reading them, because you want to know about train crashes in the last six months which is easier than doing train crashes going into Hatfield, because then you only get... only get a train crash story that is at Hatfield. So you might want to bracket off a date, possibly just to reduce the about of bumf that comes through on the machine, but I can't remember how to do that and so ...

Res So the system that you use...

LS By and large I get by.

Res You get by.

LS Yeah, I get by. I mean if I can't find out I will go back to Lexus Nexus. I generally start from the editorial database. Also when you use the editorial database it is not instantly clear where the few words are on the story. You have a 3000 word story, you can't just look through at a glance, whereas with, certainly when you do a search from my sector and I think with Lexus Nexus, it really does highlight the word properly so you can just scroll down very fast and it just flashes before your eyes and you just stop and you just scroll back slowly and pick it up again, but it doesn't work on the editorial database and the print is very slightly brighter than the print of the rest of the text, it doesn't shout at you.

Res I can see why that would be a valuable thing.

LS It is like having a very slightly off-white piece of paper amongst an entire desk of white pieces of paper, you know, if they are all strewn about, at a glance you don't see it. If you search, yeah, you will find it.

Res Really what you want to do is identify the piece of information that you are interested in based on a surface...

LS Yeah, because of a lot of the stories that you have... I mean it is only actually a few paragraphs that are going to be of any use to you.

Res And they could be quite long, you say.

LS Oh yeah, 3000 words occasionally.

Res When you have found the documents that you want to... or when you have done your searching, presuming you have identified documents which would be useful, what happens now.

LS I tend to print them out so that they are next to me. I might want to print a few, I very rarely cut and paste, partly because I am concerned about copyright, partly because I hate using other people's words when I can use my own words, and partly because it is very rare that it will actually be appropriate to cut and paste something from archives and to uncover a story.

Res So you have a pile of papers beside you that you've collected..

LS I might do.

Res When would you not have an interest, have a collection of printouts?

LS When I have no need to go to archive. That would tend to be court stories for a start.

Res Court stories?

LS Yeah, if you have got a running court story, you cannot add anything to it that hasn't been said in court in front of the jury, so all you do is colour your take on a story if you have loads of cuts and if I bollocks up and decide to insert something that you have read in cuts and you forget that its hasn't been said for the jury. I tend to avoid cuttings on court stories, except, say it is day 5, I might go back and find out what exactly was said on days 1 to 4.

Res Do you ever makes notes of information that you find?

LS From archive? I might jot something down in my notebook. I am more likely to get the printout and mark the paragraph, so that when I am looking through the pile of papers I can easily find the paragraph I am looking for. It is not failsafe.

Res How often do you go back to the documents that you have collected? Do you tend to refer to a document a number of times, or once?

LS It depends, I mean for example the story that we worked on, we were looking at yesterday was 1967 outbreak of foot and mouth disease. Now if you were told to do a story on the 1967 foot and mouth disease you would be looking at archives materials virtually 100%. That would be your primary source. If you are working on...

Say I thought a story has come to an end, the jury has convicted and the judge has sentenced, you have got all the present day's material, again you probably either have enough from agency or enough knowledge yourself of the story to fill-in the basic background of it, which you tie in with the judge's summing-up, and in run-of-the-mill court stories you wouldn't use much archive, except perhaps reference just to check what the prosecution said or just double check the charges, the specific charge or something.

On the other hand, if you are doing a more involved case, where you want the full background, you will read through every single piece of information you can get on the defendant for example, and then you cut them [unint] probably in order to do your background. Although if it is going to be accurate, the story, most of the time, you would hope that background reading worked on independently by staff, and you weren't really relying on archives. People who might know about... might be able to shed some light on the case, the defendant, the victims, whatever.

Res Right. So depending on the case, depending on the type of story, there may be instances...

LS ... the type of story in order to be organised, you need to know the theme before you sit down and write it. I mean, it may be different on a background piece on, like with regards to Jill Dando. It will be an absolute madhouse on the day he is, say he is convicted. Because then we would just be cutting archive material, cutting agency and copy type it together. What should have happened is we should have a reporter putting together a story from now, but that probably won't happen, so that all we needed really to do is just adjust it slightly according to what the judge might say, and then press the button.

Res I don't want to lead you, but would you say that it is a common experience that at some point some decisions about what you are writing might change, as you are writing and finding out information, and you then want to go back to the material, the archive material to look at some other aspect?

LS That can certainly happen. I mean there are certain... frequently instructions change or facts either change or new facts come to light that cast a different light on the story and then that means yes you may want to return to the archives.

Res Right, so it has to be quite a flexible process then?

LS You can't be inflexible as a reporter. And it should be in your job description, flexibility. It has to be, I mean there is no other way you can do the job, you have got such tight deadlines and you don't know what you are going to do from one day to the next.

Res Do you write structure plans for pieces that you write?

LS I write one very occasionally when I am on the road, I jot down 4 or 5 points I want to make in order, but no, basically I don't do structure plans.

Res And you mentioned earlier about the train crash type of story. Would it be true to say that there is a particular... there is a format at least, that is expected of stories.

LS No it is not a format.

Res It's not a format?

LS No, I mean it depends on what information you get. There are certain parameters. When you are a reporter the aim is to tell the readers what happened and so you could say that yes you've got... the first thing you want to do is have facts, but that is going to be true of every story. You know, I think we know, from what we know about our readers, what sort of things will interest them, and how you interest them and write the story. So, it works on that rather than... It is not a format, it is knowledge of what will appeal to the reader, what the reader needs to know, they want to know, would like to know. And the same [unint] Its very [unint] You make presumptions about the readers and sometimes you won't.

Res We have gone on for about 20 minutes now. Have you got time to do a quick summary of the subject matter of a breaking news story you recently wrote.

LS I am trying to think, I got a story about dormice the other day, but that was not really breaking news. I suppose it was. I don't know. I can't remember what I might have wrote. Well last time I was sent out to the office was to cover a case on industrial poisoning. On the dormouse story I did, the first thing I wanted to know off archive was has this reintroduction programme been covered by the papers before. The second thing I wanted to know was, what can I find out about Dormice. I would like to have a purpose, what could I have it on. I might have gone further and said what I know about the people who are doing this reintroduction programme.

On the industrial point, in that one, the first thing I would want to do is check whether that particular company, and I have got a company name, so [unint]. Second thing I would have wanted was a generalised... I would like some information on the chemical itself would be useful. What are its dangers, what are they used for. And I would also have wanted to be able to find out recent examples... not even recent, just past examples of chemical spillages that had actually caused injury, whether it be that particular chemical which

was something tetrachloride. If the story had turned out as it was initially billed, it went rather flat on its face in the end, but if there had been fifty people who were injured, I needed to know examples of death, who has collapsed? And be able to put that in the big piece and get some background details on previous cases as well. Those who were in the first [unint] never wanted [unint] I mean I can't actually think of... certainly when I do it, when I am accessing our current system and I have to try and think of the relevant keywords and sometimes it is very difficult to think of a narrow enough keyword.

Res Because you just get reams and reams of information out.

LS Well for example on this story I wouldn't have necessarily have mentioned this tetrachloride stuff if I had done a previous story. It might have been not found. On the other hand, if I put chemicals in as a keyword, I would have got thousands of stories, so thinking of some way to narrow the thing. You don't want to be too narrow. 'Industrial accident', hopefully that would have done the trick.

Res Do you narrow by putting in more keywords?

LS Yeah, I like to be able to, but it doesn't really work. And also you can't, because there isn't an automatic keyword, there might be two or three keywords, you throw up from the first keyword. You won't bother with the second, you are wasting your time, but it may happen to be that the second or third keyword answers precisely the question you were looking for, as opposed to roughly.

Res So you might be cycling a few times to get the query as you want it?

LS I can't honestly ensure an easy way round it.

Res Do you use biographies at all?

LS Whether books or...? I tend to use things like Who's Who, you know and I would often prefer them to be in far more detail, whether they be online or paper. And then you haven't got the library files.

Res Someone has done a background?

LS Yeah, but I think the library hasn't actually been useful to me yet. I think the library occasionally will fix on somebody who is in the news quite a lot, and therefore they will gather some details and put them on the files. If you do that for every individual that would be even better.

Res Yeah

LS But that isn't practical.

Res No it is a very intensive thing. Perhaps I was thinking about the possibility of getting a machine to do it automatically, but that is very complicated.

LS Yes, and I am not quite sure how you could do it, how you can teach your machine to recognise the details that you are looking for, they might vary from day to day.

Res Do you find yourself looking for quotations?

LS On the day, definitely and sometimes from archives. If I need wordings of what somebody said or... I can't think of a precise example, but there are occasions when you know somebody has said something important on the subject and you know that there is this particular quote. Its a bit like being told in the Ancient Mariner, some people think it is 'not a drop to drink', 'nor a drop to drink', I think it is actually 'nor any drop to drink' and you would probably need to check if you were going to quote that line. You would need to check which were right rather than what you think you can remember correctly.

Res How would you find that?

LS That one, erm.

Res A copy of the Ancient Mariner?

LS I would get a copy of the Ancient Mariner, yeah. Somewhere over there [indicates]

Res Okay that is fantastic, that has been enormously valuable.

LS OK.

Appendix IIIc

Cat. 1 Constraints

< Properties and dimensions >
Prescribes:
Determinant(s):

< Subcategories>
Cause of any a priori indeterminability:

Cat. 1.1 Angle

< Properties and dimensions >
Prescription: That a specified report follow a specified angle.
Determinant(s): Angle.

< Subcategories>
Source(s): Editorial decision making.
Cause of any a priori indeterminability: Commitments at the paper level and knowledge of what is being reported can evolve and be revised throughout an assignment. Misunderstandings can also occur concerning the intended angle which are subsequently resolved.

Cat. 1.2 Deadline

< Properties and dimensions >
Prescription: That a specified report be finished by a specified time.
Determinant(s): Deadline.

< Subcategories>
Source(s): Editorial decision making.
Cause of any a priori indeterminability: Decisions about which page to put a report (which determines the deadline) can evolve and can be revised throughout an assignment.

Cat. 1.3 Word count

< Properties and dimensions >
Prescription: That a specified report consist of a specified number of words.
Determinant(s): Word count.

< Subcategories>
Source(s): Editorial decision making.
Cause of any a priori indeterminability: Commitments at the paper level evolve and can be revised throughout an assignment.

Cat. 1.4 Prior written commitments

< Properties and dimensions >
Prescribes: That subsequently written test coheres with previously written text.
Determinant(s): Previously written text.

< Subcategories>
Source(s): The writer.

Cause of any a priori indeterminability: Commitments at the report level evolve and potential revision throughout an assignment.

Cat. 1.5 Constraints on content

< Properties and dimensions >
Prescribes: Aspects of the content of a report.
Determinant(s):

< Subcategories>
Source(s):
Cause of any a priori indeterminability:

Cat. 1.5.1 Newsworthiness constraints

< Properties and dimensions >
Prescribes: That the content of a report be newsworthy.
Determinant(s):

< Subcategories>
Source(s):
Cause of any a priori indeterminability:

Cat. 1.5.1.1. Timeliness/Currency constraint

< Properties and dimensions >
Prescribes: That the content of a report be concerned with recent event(s).
Determinant(s): Recent events.

< Subcategories>
Source: Many
Cause of any a priori indeterminability:

Cat. 1.5.1.2. Proximity constraint

< Properties and dimensions >
Prescribes: That the content of a report prioritise issues local to reader.
Determinant(s): The readership and the locality of aspects of what is being reported

< Subcategories>
Source: A model of the readership and what is being reported.
Cause of any a priori indeterminability: Knowledge of what is being reported can evolve throughout an assignment.

Cat. 1.5.1.3. Exclusivity constraint

< Properties and dimensions >
Prescribes: That the angle of a report be different from pieces published before in prominent national newspapers and magazines.
Determinant(s): The angles of pieces published before in prominent national newspapers and magazines recently.

< Subcategories>
Source: Cuttings archive.
Cause of any a priori indeterminability: The journalist's knowledge of the angles of what was published before in prominent national newspapers and magazines recently can evolve during an assignment.

Cat. 1.5.1.4. Human interest constraint

< Properties and dimensions >
Prescribes: That the content of a report prioritises human interest.
Determinant(s): Aspects of what is being reported with high human interest.

< Subcategories>
Source(s): (Depends on context)
Degree of a priori indeterminability: Medium
Cause of indeterminability: Knowledge of aspects of what is being reported with high human interest can evolve throughout an assignment.

Cat. 1.5.2 Historical context constraint

< Properties and dimensions >
Prescribes: That the report relates what is being reported to relevant historical context.
Determinant(s): What is being reported and relevant historical context.

< Subcategories>
Source(s): (Depends on context)
Cause of any priori indeterminability: Knowledge of relevant historical context can evolve throughout an assignment.

Cat. 1.5.3 Accuracy constraint

< Properties and dimensions >
Prescribes: That within content of a report propositional claims be true, and name spellings be correct.
Determinant(s): The facts and correct spellings.

< Subcategories>
Source(s): (Depends on context)
Cause of any a priori indeterminability: Knowledge of what is true and name spellings can evolve throughout an assignment.

Cat. 1.5.4 Legal constraints

< Properties and dimensions >
Prescribes: That the content of a report should not transgress any laws.
Determinant(s): The law.

< Subcategories>
Source(s): Government.
Cause of any a priori indeterminability:

Cat. 1.5.4.1. Libel constraint

< Properties and dimensions >
Prescribes: That the content of a report should not transgress libel law.
Determinant(s): The law of libel.

< Subcategories>
Source(s): Government.
Cause of any a priori indeterminability:

Cat. 1.5.4.2. Perjury constraint

< Properties and dimensions >
Prescribes: That the content of a report should not transgress perjury law.
Determinant(s): The law of perjury.

< Subcategories>
Source(s): Government.
Cause of any a priori indeterminability:

Cat. 1.5.5 Explanation constraint

< Properties and dimensions >
Prescribes: That the content of a report should explain reported events.
Determinant(s): The cause of reported events

< Subcategories>
Source(s): Various.
Cause of any a priori indeterminability: Knowledge of causes can evolve throughout an assignment.

Cat. 1.6 Constraints on structure

< Properties and dimensions >
Prescribes: That a news report or feature should be structured in certain ways.
Determinant(s):

< Subcategories>
Cause of any a priori indeterminability:

Cat. 1.6.1 Cut-from-bottom constraint

< Properties and dimensions >
Prescribes: That a news report should be structured such that information appears in order of importance with the most important information first.
Determinant(s): The importance of different pieces of information.

< Subcategories>
Source(s): The story being reported.
Cause of any a priori indeterminability: Decisions about what information to report can evolve with knowledge of the story.

Cat. 1.6.2 Original wording constraint

< Properties and dimensions >
Prescribes: That a news report of feature should be structured differently at the word level from reports published before.
Determinant(s): The wording of pieces published before.

< Subcategories>
Source(s): Written news sources.
Cause of any a priori indeterminability: The journalist's word level knowledge of what was published before can evolve during an assignment.

Cat. 2 Information Resources

< Properties and dimensions >>
Location: External/internal to the journalist
Mode of access: Read-only ... read-write.
Interaction paradigm supported: Passive ... active.

< Subcategories>

Cat. 2.1 External information resources

< Properties and dimensions >
Location: External to the journalist.
Mode of information access: Read-only ... read-write
Interaction paradigm supported: Passive ... active
Information Scope:
Mobility:

< Subcategories>

Cat. 2.1.1 External read-only information resources

< Properties and dimensions >
Location: External to the journalist
Mode of access: Read-only
Interaction paradigm supported: Passive ... active.
Information Scope:

Cat. 2.1.1.1. Electronic news cuttings (ENC) service

< Properties and dimensions >
Location: External to the journalist
Mode of access: Read-only.
Interaction paradigm supported: Active.
Information Scope: Past news stories.

Cat. 2.1.1.2. News Library

< Properties and dimensions >
Location: External to the journalist
Mode of access: Read-only.
Interaction paradigm supported: Active.
Information Scope: Past news stories.

Cat. 2.1.1.3. Newswire resources

< Properties and dimensions >
Location: External to the journalist.
Mode of access: Read-only.
Interaction paradigm supported: Passive ... active.
Information Scope: Breaking news summaries.

Cat. 2.1.1.3.1. Copy Taster alerts

< Properties and dimensions >
Location: External to the journalist.
Mode of access: Read-only.
Interaction paradigm supported: Passive.
Information Scope: Breaking news summaries on a given topic.

Cat. 2.1.1.3.2. Newswire archive

< Properties and dimensions >
Location: External to the journalist.
Mode of access: Read-only.
Interaction paradigm supported: Active.
Information Scope: Past breaking news summaries.

Cat. 2.1.1.4. Informants

< Properties and dimensions >
Location: External to the journalist.
Mode of access: Read-only.
Interaction paradigm supported: Passive ... active.
Information Scope:

Cat. 2.1.1.4.1. Witnesses

< Properties and dimensions >
Location: External to the journalist.
Mode of access: Read-only.
Interaction paradigm supported: Active.
Information Scope: A specific event from a specific perspective.

Cat. 2.1.1.4.2. Experts

< Properties and dimensions >
Location: External to the journalist.
Mode of access: Read-only.
Interaction paradigm supported: Active.
Information Scope: Detailed information and informed opinion on a specific topic.

Cat. 2.1.1.4.3. Colleagues

< Properties and dimensions >
Location: External to the journalist
Mode of access: Read-only
Interaction paradigm supported: Active
Information Scope: Various

Cat. 2.1.2 External read-write information resources

< Properties and dimensions >
Location: External to the journalist.
Mode of access: Read-write.
Interaction paradigm supported: Active.
Information Scope:

Cat. 2.1.2.1. The 'holding document'

< Properties and dimensions >
Location: External to the journalist.
Mode of access: Read-write.
Interaction paradigm supported: Active.
Information Scope: Information extracts considered potentially useful at some time during the research and writing process.

Cat. 2.1.2.2.Printouts

< Properties and dimensions >
Location: External to the journalist.
Mode of access: Read-write.
Interaction paradigm supported: Active.
Information Scope: Documents and information considered potentially useful at some time during the research and writing process.

Cat. 2.1.2.3.Written report structure plans

< Properties and dimensions >
Location: External to the journalist.
Mode of access: Read-write.
Interaction paradigm supported: Active.
Information Scope: Report structure decisions.

Cat. 2.2 Internal information resources (knowledge)

< Properties and dimensions >
Location: Internal to the journalist.
Mode of access: Read-write.
Interaction paradigm supported: Passive ... active
Information Scope:

Cat. 2.2.1 Domain knowledge

< Properties and dimensions >
Location: Internal to the journalist.
Mode of access: Read-write
Interaction paradigm supported: Passive ... active
Information Scope: A more or less partial knowledge of the domain.

Cat. 2.2.2 General writing knowledge

< Properties and dimensions >
Location: Internal to the journalist.
Mode of access: Read-write
Interaction paradigm supported: Passive ... active
Information Scope: A more or less partial knowledge of writing conventions and methods.

Cat. 2.2.3 Newspaper writing knowledge

< Properties and dimensions >
Location: Internal to the journalist.
Mode of access: Read-write
Interaction paradigm supported: Passive ... active
Information Scope: A more or less partial knowledge of news reporting conventions and methods.

Cat. 2.2.4 Resource knowledge

< Properties and dimensions >
Location: Internal to the journalist.
Mode of access: Read-write
Interaction paradigm supported: Passive ... active
Information Scope: A more or less partial knowledge of tool operating methods.

Cat. 3 Information behaviours

Cat. 3.1 Information Seeking

< Properties and dimensions >
Goal: Finding information
Target: Information
Focus: Low ... medium ... high
Extent of searching: Light ... thorough.

< Subcategories>
Rationale:
Circumstances:
Means:
Consequence:
Desired means:

Cat. 3.1.1 Exclusivity Checking

< Properties and dimensions >
Goal: Finding information to confirm
Target: Previously published angles similar to the current angle.
Focus: Medium
Extent of searching: Light...thorough

< Subcategories>
Rationale: Originality (+ finding useful information)
Circumstances: At the beginning of all assignments except for diary jobs where they are already sure or they are reporting scientific research (where originality is assumed)
Means: Online archive, library, looking for reports published in: the previous two weeks, national newspapers and prominent magazines (not local or international) external contacts
Consequence: Potential angle rejects or modification (traded-off against newsworthiness)
Desired means:

Cat. 3.1.2 Background Information seeking

< Properties and dimensions >
Goal: Finding information.
Target: Background information on a specified subject.
Focus: Low ... medium ... high
Timing: Beginning ... middle ... end
Extent of searching: Light ... thorough.

< Subcategories>
Rationale: To understand the subject in a historical context. To describe events in a historical context.
Circumstances: When writing a story with history. More with features then news stories. Always.
Means: Online cuttings search, delegating search to library, contacting outside agencies, looking through personal collection.
Consequence:
Desired means:

Cat. 3.1.2.1. Seeking background overviews

< Properties and dimensions >
Goal: Finding information
Target: Background information on a specified subject in the form of an overview or summary.
Focus: Low
Timing: Beginning
Extent of searching : Light...thorough
Frequency: High.
Objective: Discover ... remind ... confirm

< Subcategories>
Rationale: To understand the subject in a historical context. To describe events in a historical context. Knowing what to ask in an interview. To reduce research time. Comprehensiveness.
Circumstances: When writing a story with history. More with features then news stories. Always. When domain knowledge is low (novices or reporters working off patch), before an interview. When writing a backgrounder.
Means: Online cuttings search for a big piece or a few features or a piece with an explicit chronology, or pieces in broadsheets or American papers, getting a library 'fact-file' or delegating cuttings search to library, newswire search, contacting outside agencies, looking through personal collection of information. By asking an informed colleague.
Consequence:
Desired means:

Cat. 3.1.3 Seeking evidence for a hypothesis

< Properties and dimensions >
Goal: Finding information
Target: Information which supports or falsifies a proposition.
Focus: Low ... medium ... high
Timing: Beginning ... middle ... end
Extent of searching: Light ... thorough.
Objective: Discover ... remind ... confirm

< Subcategories>
Rationale: To understand the subject in a historical context. To describe events in a historical context.
Circumstances: When writing a story with history. More with features then news stories. Always. When the angle is a hypothesis.
Means: Online cuttings search and sifting for facts or anecdotes or examples. Delegating search to library. Contacting outside agencies. Looking through personal collection.
Consequence:
Desired means:
Frequency: High

Cat. 3.1.4 Information seeking for feature comparison

< Properties and dimensions >
Goal: Finding information for comparison.
Target: Facts and figures
Focus: High
Timing: Beginning ... middle ... end
Extent of searching: Light ... thorough.
Objective: Discover ... remind ... confirm

< Subcategories>
Rationale: To make comparison with current news. To understand the subject in a historical context. To describe events in a historical context.
Circumstances: When writing a story with history. More with features then news stories. Always.
Means: Online cuttings search, delegating search to library, contacting outside agencies, looking through personal collection.
Consequence:
Desired means:
Frequency: High ... low

Cat. 3.1.4.1. Seeking properties of past disasters

< Properties and dimensions >

Goal: Finding information for comparison.

Target: The cause, location, numbers of dead or injured and subsequent prevention measures relating to a specified disaster.

Focus: High

Timing: Beginning ... middle ... end

Extent of searching: Light ... thorough.

Objective: Discover ... remind ... confirm

< Subcategories>

Rationale: To make comparison with a current disaster. To establish if prevention measures were adequate. To establish patterns of common causation or geography. To look for potential explanations for current disasters. Because readers will be interested. To compare severity (numbers of dead and injured). To establish ways in which a current disaster is remarkable. To understand the subject in a historical context. To describe events in a historical context.

Circumstances: When writing about a disaster or catastrophe. When writing a story with history. More with features then news stories. Always.

Means: Online cuttings search, delegating search to library, contacting outside agencies, looking through personal collection.

Consequence:

Desired means:

Frequency: High ... low

Cat. 3.1.4.2. Discovering/confirming what someone said

< Properties and dimensions >

Goal: Finding information for comparison.

Target: What was said, or a specified thing said, by a specified person about a specified subject either within or not within a specific time-frame.

Focus: High

Timing: Beginning ... middle ... end

Extent of searching: Light...thorough

Frequency: High

Objective: Discover/confirm

< Subcategories>

Rationale: To find inconsistencies. Correspondence. To avoid misquoting (which is common). To understand the subject in a historical context. To describe events in a historical context.

Circumstances: When writing a political story. When writing about an on-going court case. When writing a story with history. More with features then news stories. Always.

Means: Online cuttings search, using person's name as keywords, and expected words from the quote delegating search to library, contacting outside agencies, looking through personal collection.

Consequence:

Desired means: Dedicated quotation search tool incorporating subject keyword and speaker search tool which shows the quote in its report context.

Cat. 3.1.5 Confirming names and how to spell them

< Properties and dimensions >
Goal: Finding information to confirm.
Target: The name of a specified entity
Focus: Very high
Timing: Beginning ... middle ... end
Extent of searching: Light ... thorough
Frequency: High
Objective: Confirm

< Subcategories>
Rationale: Correspondence.
Circumstances:.
Means: Online cuttings search (but often cuttings get it wrong too), comparing frequencies of different spellings on the Web, contacting outside agencies, looking through personal collection.
Consequence:
Desired means:

Cat. 3.1.6 Identifying useful contacts

< Properties and dimensions >
Goal: Finding information.
Target: The name of an external agent fulfilling a specified criteria who might be able to provide information.
Focus: High
Timing: Beginning ... middle ... end
Extent of searching: Light ... thorough.
Objective: Discover

< Subcategories>
Rationale:
Circumstances: When writing a feature about a technical subject such as research or social affairs or a big crime case or disaster. When an expert opinion is needed or victims view.
Means: By searching online cuttings for expert comment and assessing the quality, opportunistically during reading, asking colleagues, press offices, contacting experts in related fields, searching company websites
Consequence:
Desired means:

Cat. 3.2 Information Gathering

< Properties and dimensions >
Goal: Gathering information.
Source:
Destination:

< Subcategories>
Rationale: To enable fast relocation of information later (without complete re-reading) to be reminded of the facts and figures (which are easy to forget) such as ages, details of jobs and careers, locations, and perhaps 'weave' them into a report to add background or colour, or perhaps to illustrate a point. To provide a coverage checklist. To indicate a line to follow up to focus further investigation, perhaps in an interview.
Circumstances: When information is found in a document that is considered useful or is potentially useful, such as a good quotation or an explicit chronology.
Means:
Consequence:
Desired means:

Cat. 3.2.1 Dragging and dropping

< Properties and dimensions >
Goal: Gathering information.
Source: On-screen text files including cuttings and newswires
Destination: On-screen text editor file or 'basket', 'work-paste-pad', 'personal database' etc. or copy directly into copy.

< Subcategories>
Rationale: To enable fast relocation of information later (without complete re-reading) to be reminded of the facts and figures (which are easy to forget) such as ages, details of jobs and careers, locations, and perhaps 'weave' them into a report to add background or colour, or perhaps to illustrate a point.To provide a coverage checklist. To indicate a line to follow up to focus further investigation, perhaps in an interview. Because it's easy using a split screen.
Circumstances: When reading onscreen. When time is short. When you don't have to change text, or only a bit. When there is no mobility requirement. After skim reading the printed version. When information is found about a potential contact. When a little information is found in a document that is considered useful or potentially useful, such as a good quotation or an explicit chronology.
Means:
Consequence: Can lead to unintentional plagiarism.
Desired means:

Cat. 3.2.2 Printing

< Properties and dimensions >
Goal: Gathering information.
Source: On-screen text files including cuttings and newswires.
Destination: Paper.

< Subcategories>
Rationale: To enable fast relocation of information later (without complete re-reading) to be reminded of the facts and figures (which are easy to forget) such as ages, details of jobs and careers, locations, and perhaps 'weave' them into a report to add background or colour, or perhaps to illustrate a point. To provide a coverage checklist. To indicate a line to follow up to focus further investigation, perhaps in an interview. So they can move the piece around the office to show other people or take out of the office. So they can easily relocate any of the information later (without searching again). To support the substantiating of facts in case of a subsequent challenge such as from the editor, a complaint or a legal challenge. Relocating information in a printout is quicker than it is onscreen. Reading a printout is more pleasant than reading onscreen and easier to skim.
Circumstances: When a complex cutting is found. When time is not short. Most feature and some news assignments. When there is a mobility requirement. When a lot of information is found in a document that is considered useful or potentially useful such as a good quotation or an explicit chronology. When useful information is found during the skim reading of a printed cutting.
Means:
Consequence: Can avoid unintentional plagarism.
Desired means:

Cat. 3.3 Information Reviewing

< Properties and dimensions >
Goal: Relocating information.
Target: Information which have already been found during the course of an assignment.
Subject: ....
Focus: Low ... medium ... high
Timing: Beginning ... middle ... end
Extent of searching: Light ... thorough.

< Subcategories>
Rationale:
Circumstances:
Means:
Consequence:
Desired means:

Cat. 3.3.1 Reviewing information gathered during an assignment

< Properties and dimensions >
Goal: Relocating information.
Target: Specific items of information which have already been found during the course of an assignment and which have been placed in an assignment specific collection.
Focus: Low ... medium ... high
Timing: Beginning ... middle ... end
Extent of searching: Light ... thorough.

< Subcategories>
Rationale:
Circumstances:
Means: Spacial location, visual recognition of highlighting
Consequence:
Desired means:

Cat. 3.3.2 Reviewing information read but not gathered during an assignment

< Properties and dimensions >
Goal: Relocating information.
Target: Information which have already been found during the course of an assignment but which have not been placed in an assignment specific collection.
Focus: Low ... medium ... high
Timing: Beginning ... middle ... end
Extent of searching: Light ... thorough.

< Subcategories>
Rationale: To flesh something out.
Circumstances: When they failed to see the significance of some information because they didn't have the full grasp (254). Or they omitted to collect the information – (usually the date or name of publication) (JV 8/6)
Means:
Consequence: Frustration (13)
Desired means:

Cat. 3.3.3 Reviewing information read prior to an assignment

< Properties and dimensions >
Goal: Relocating information.
Target: Information which has been found prior to an assignment.
Focus: Low ... medium ... high
Timing: Beginning ... middle ... end
Extent of searching: Light ... thorough.

< Subcategories>
Rationale:
Circumstances:
Means:

Appendix IV

Appendix IVa

Task 1

1st newswire

Bob Taylor [founder of the successful BNS news corporation][1] has been rushed to hospital after being attacked by burglars he disturbed at his 120,000-acre ranch just outside of Bozeman, Montana early yesterday morning. Taylor was rushed to Bozeman Deaconess Hospital with a fractured skull. A hospital spokesman said that Taylor is in a serious but stable condition.

1st brief

You work for a British broadsheet newspaper. Write a 200 word report to appear as a single column story with the angle, 'The latest in a line of misfortune for Bob. Will he bounce back?'

Starter query: BNS

2nd newswire

A hospital spokesman at Bozeman Deaconess Hospital where Bob Taylor was being treated has announced that Taylor died suddenly from a brain hemorrhage at around 9 O'clock today. They say that Taylor hadn't regained consciousness since he had been admitted early this morning.

2nd brief

A correspondent in the US will now write the main story. What we need from you is a 300 word backgrounder to go alongside this in a sidebar. The angle is to be: 'The rise and fall of Bob Taylor'.

Task 2

1st newswire

Two journalists, American Robert Walsh and Briton James Whittle, were kidnapped yesterday close to the southern Colombian town of Mitu. It is believed they were taken by the Marxist guerrilla group, the Revolutionary Armed Forces of Colombia (FARC) after they published photographs of a FARC-run poppy field and named its location.

1st brief

Write a 200 word backgrounder to go alongside the main story with the angle: 'This is the latest in a series of journalist kidnappings in Columbia'.

Starter query: Kidnapped

2nd newswire

Journalists Robert Walsh and James Whittle have been found dead by the Colombian police by a roadside in a mountainous region outside the south-

[1] This note is for the reader's information, and was not included in the newswires handed to subjects.

western city of Cali where it was believed they were being held. The police say that both journalists had been shot several times and appeared to have died instantly. FARC guerrillas are suspected, but this has not yet been confirmed.

2nd brief

We now need a 300 word backgrounder with the angle: 'This is the latest in a spate of journalist killings in Colombia'.

Task 3

1st newswire

A Russian news agency has reported that journalist Mishka Yanko, who had been serving a four-year sentence in a Russian prison for treason until he was released on parole last month, has committed suicide. Russian police say that three days ago Yanko and his wife disappeared from the house where he was being kept under surveillance.

1st brief

Write a 200 word backgrounder with the angle, 'Yanko is dead after years of victimization by the Russian authorities'.

Starter query: Russian journalist

2nd newswire

It turns out that Yanko isn't dead after all. He had broken his parole and gone to France helped by the press rights group Reporters Without Borders.

2nd brief

We now need a 300 word backgrounder with the angle: 'Yanko escapes years of victimization by the Russian authorities'.

Appendix IVb

Subject Instructions

Common

In this session you will be asked to write (a/another) news story based on a newswire report, some editorial instructions, and information that you find by doing searches on a news cuttings archive. The task has been designed to be as real-to-life as possible.

You will be working to a 50min deadline and the screen as well as what we say will be recorded. After this you will be asked to fill in a questionnaire giving some opinions about the software and equipment you used.

The system you see on the screen has a search engine which searches past news stories. All the news stories are about media issues. The query box [indicate] allows you to enter search terms. When you then press search, the system will look for news stories containing these terms. For example, you can enter the word 'Bush', and the system will search for stories containing that word, which might include 'President Bush' as well as phrases such as 'behind the bush'. The news stories are listed in date order with the latest at the top. You can type more than one word, for example 'Bush Blair' and the system will search for stories that contain both words – but not necessarily next to each other. The search engine will not allow you to specify more complex searches, such as phrases, documents that contain a particular word or another, or documents that don't contain a particular word.

Before you start the task a search that may be useful will have been run to get you started.

When you submit a search, the headlines of the matching documents are shown in the window below. You can then click on any of these to see the full text in the middle window.

Condition NHP

If you find material in a news story that you would like to retain you can print the story by clicking the print button. The printout will appear with a large number on it – you can ignore this. You are provided with a highlighter pen so you can mark pieces of text to make the easier to find later.

The right-hand window is where you can write notes and where I would like you to write your news report – but please don't use this area to reproduce information from the articles you read verbatim – use printing for this.

In the right hand window you can reformat text as a heading or as ordinary text by highlighting it and clicking the relevant formatting button above.

Important: Using the number on the printouts, I will note-down the printed documents you read as you work. I would also like to note-down when you read

text that you haven't highlighted. When you refer to a document and read some text that is not highlighted, please say the first word you read out loud. Only do this once for each time you turn your eyes to the document. So if you then look at the screen or at another document, and return to the original document, please, once again, say the first word of any un-highlighted text you read. Do this for all the documents you read.

Do you have any questions?

Condition NHR

If you find material in a news story you would like to retain, you can highlight it with the mouse and drag it into the right hand window where it will be marked as a quotation by appearing indented, in dark red text. Please don't use this area to reproduce information from the articles you read verbatim.

The right-hand window is also the place where you can type your own notes, and where I would like you to write your report. Text you type will not be indented and will appear in black.

You can format text in the right window as a heading by highlighting it and clicking the "heading" button. There are also buttons for changing text to notes format or quotation format in the same way. Any piece of text can be formatted as any of the three styles.

Do you have any questions?

Condition NHO

If you find material in a news story you would like to retain, you can highlight it with the mouse and drag it into the right hand window where it will be marked as a quotation by appearing indented, in dark red text. – but please don't use this area to reproduce information from the articles you read verbatim.

Any text dragged into the right-hand window will have a link added to the end. You can click this link if you want the document it came from to appear in the middle window again. When you do this, the text you originally dragged will be highlighted in the middle window in yellow. Any other text you have dragged from a document will also be highlighted in grey.

The right-hand window is also the place where you can type your own notes, and where I would like you to write your report. Text you type will not be indented and will appear in black.

You can format text in the right window as a heading by highlighting it and clicking the "heading" button. There are also buttons for changing text to notes format or quotation format in the same way. Any piece of text can be formatted as any of the three styles.

Do you have any questions?

Appendix IVc

Post-task questionnaire

Subject name..... Subject number .....

Please mark a point on each line indicated by the experimenter that best indicates your agreement with the statement or tick the NA box.

If you have just completed the second or third session of the study, you may wish to revise your previous ratings – please feel free to do this.
the statement

1. During the task, I found I wanted to re-consult source documents to find information I had remembered reading but did not necessarily consider useful at the time.

NHO Strongly disagre ——— Strongly agree NA☐

NHR Strongly disagre ——— Strongly agree NA☐

NHP Strongly disagre ——— Strongly agree NA☐

2. This happened because later I found wanted to include specific items of information to my report.

NHO Strongly disagre ——— Strongly agree NA☐

NHR Strongly disagre ——— Strongly agree NA☐

NHP Strongly disagre ——— Strongly agree NA☐

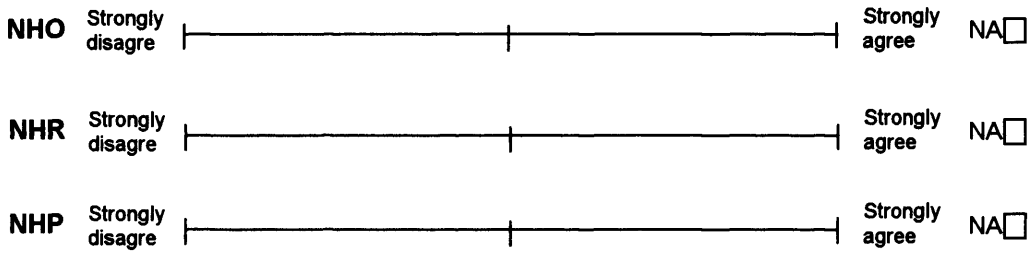
3. This happened because later I wanted to better understand the context of information which I **had** identified as useful.

NHO Strongly disagre ——— Strongly agree NA☐

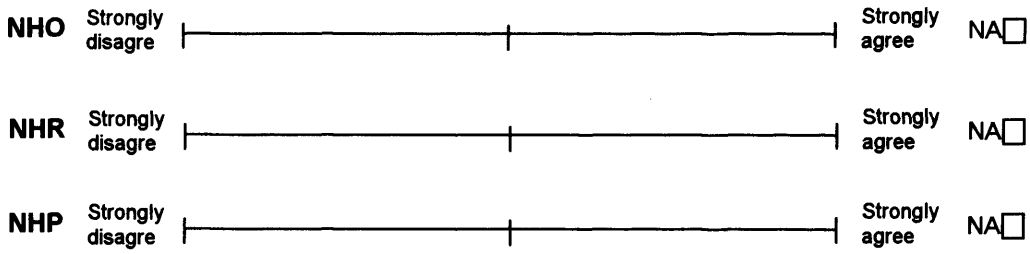
NHR Strongly disagre ——— Strongly agree NA☐

NHP Strongly disagre ——— Strongly agree NA☐

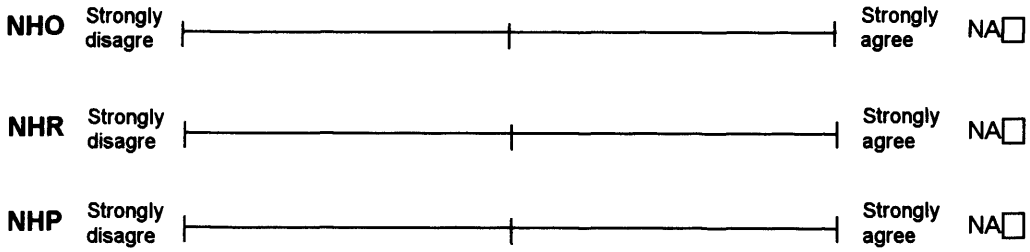
4. This happened because later I wanted to re-consult source documents to check if there was anything else I could add to my report.



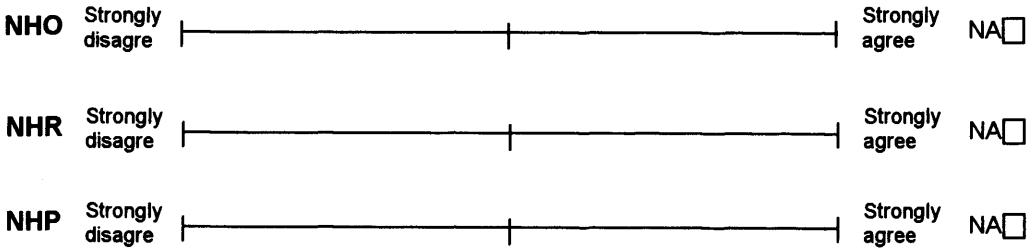
5. During the task, the actions I performed to ensure that I would be able to find useful information later took very little time.



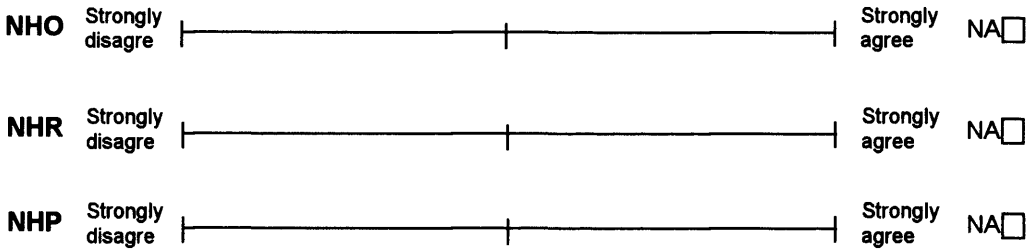
6. During the task, the actions I performed to relocate this information took very little time.



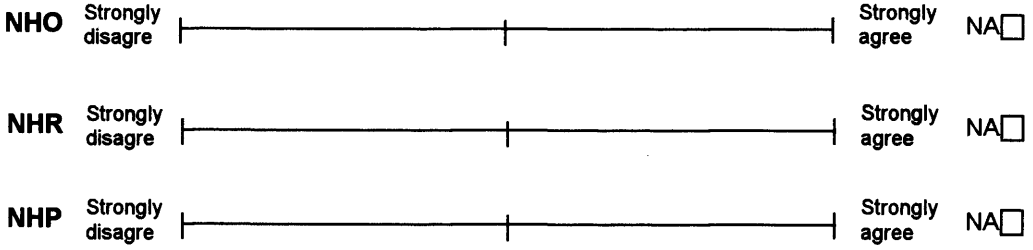
7. Relocating information that I had not initially identified as useful, but which <u>was</u> in a document containing other information that I had identified as useful, took very little time.



8. The set-up I just used allowed me to work in what I regard as a flexible and dynamic way.



9. I enjoyed using the set-up.



Appendix IVd

Figure IVc1 (adapted from Senn, 1997, pp209) is used to demonstrate the logic underlying equivalence and non-inferiority testing. In the figure, the horizontal line at the bottom is an axis on which values representing the difference between the population means for two conditions, or treatments can be located. By convention, in equality and non-inferiority trials, this difference is represented as: $U_E - U_R$, where U_E is the population mean for the experimental condition, and U_R is the population mean for the reference condition. Hence, a positive difference indicates $U_E > U_R$, a negative difference indicates $U_E < U_R$, and zero difference indicates $U_E = U_R$. In figure IVc1, zero difference is represented by the vertical dotted line.

In these terms, it might be considered that the goal of equality testing, ought to be to demonstrate that $U_E = U_R$ i.e. that the difference between the population means is zero. However, demonstrating exact equality of population means on the basis of samples of those populations is, in principle, not possible (Senn, 1997 pp212). The best that can be hoped for using population samples is to calculate the probability that the difference between two population means falls within a specified range of values (i.e. a confidence interval).

Given this possibility, the approach adopted by equality testing is first to define a range each side of exact equality which is sufficiently small that, should the difference between two populations fall within it, for all intents and purposes they can be considered equivalent; this is termed a *region of practical equivalence* (Senn, 1997 pp 208). The extent of the region of practical equivalence is, of course, arbitrary, although in drugs trials it is typically set at ±20% of the mean of the reference sample. In the tests in this study slightly more rigid boundaries will be set at ±15% of the mean of the reference sample. In figure IVc1, a region of practical equivalence falls between the two vertical lines marked $-\delta$ and δ.

The question, then, for equivalence testing, is whether or not the confidence interval for the difference between two populations falls entirely within the region of practical equivalence. If it does, then it can be claimed that, at the level of confidence associated with the confidence interval, the difference between the two population means is within the region of practical equivalence and they can therefore be considered equivalent. Further, if higher scores show improved performance, the question for non-inferiority testing is whether the lower confidence limit lies above the lower bounds of the region of equivalence. If it does, then it can be concluded (within the assumptions of the method) that the difference between the population means is above this level, and therefore that the experimental condition is at least as good or better than (i.e. not-inferior to) the reference condition. Effectively, this carries the same logic as classical superiority testing in which it must be shown that the lower confidence limit is above zero difference, except that the threshold for demonstrating significance is lower by 20% of the mean of the reference sample. Hence it represents a more relaxed requirement.

To demonstrate these points, figure IVc1 shows four example point estimates of difference with associated confidence intervals (A to D) superimposed over the region of practical equivalence.

Example A represents a situation where the difference estimate falls below zero (and hence that the U_E is lower than U_R). However, the confidence interval is small and does not extend beyond the upper or lower limits of the region of practical equivalence. Hence, in the case of an equivalence test, the two population means would be considered to be effectively equivalent. Non-inferiority is also demonstrated. In this case it is only the lower

limit that is important. As an aside, since the confidence interval cuts zero difference, on a conventional difference test, the difference would be regarded as non-significant.

Example B is a case where the point estimate of the difference between populations is exactly zero. However, since the upper and lower confidence limits fall outside the region of practical equivalence it cannot be concluded that the populations are equivalent. At the level of certainty associated with the confidence interval, the actual population difference might well lie above or below the region of practical equivalence. And since it might lie below the region, non-inferiority is similarly not demonstrated.

In example C, the point estimate of the difference actually falls above the zero difference line (i.e. U_E is greater than U_R) and yet, since the upper confidence limit is outside the region, equivalence is not demonstrated. However, since the lower confidence limit is above the lower limit of the boundary, non-inferiority *is* demonstrated.

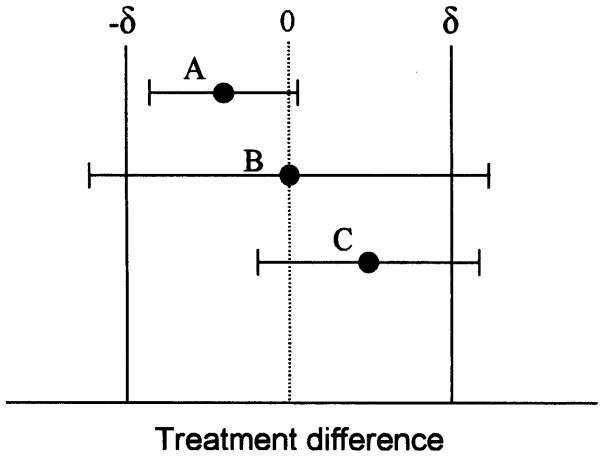


Figure IVc1. Three example point estimates and associated confidence intervals superimposed over a region of practical equivalence